现代数学译丛 4

最优化问题的扰动分析

〔法〕J.F.博南
〔美〕A.夏皮罗 著

张立卫 译



现代数学译丛 4

最优化问题的扰动分析

〔法〕J.F. 博南
〔美〕A. 夏皮罗 著

张立卫 译

科学出版社
北京

图字: 01-2007-3536 号

内 容 简 介

本书是最优化领域关于最优化问题的解如何依赖于参数扰动而变化，以及相关的一阶尤其是二阶最优性条件的最新成果的专著. 作者把很多在当前文献中不太常见的素材综合在一起，形成一完整的理论体系. 本书给出了凸分析、对偶理论等有价值的若干专题的丰富素材，很多素材在其他文献中没有出现过. 本书还详细地研究了最优化问题扰动理论在非线性半定规划和非线性半无限规划中的应用. 尤其，本书既讨论了无穷维的优化问题，又讨论了有穷维的优化问题.

本书可供运筹学与控制论专业的研究生及从事相关学科研究的研究人员参考.

Translation from the English language edition:
Perturbation Analysis of Optimization Problem
by J.Frederic Bonnans and Alexander Shapiro

图书在版编目(CIP)数据

最优化问题的扰动分析/〔法〕J. F. 博南,〔美〕A. 夏皮罗著. 张立卫 译.
—北京: 科学出版社, 2008
(现代数学译丛; 4)
ISBN 978-7-03-020429-5
I. 最… II. ①J… ②A… ③张… III. 最佳化–扰动–分析 IV. O224
中国版本图书馆 CIP 数据核字 (2008) 第 064068 号

责任编辑: 赵彦超 / 责任校对: 陈玉凤
责任印制: 赵德静 / 封面设计: 王 浩

科 学 出 版 社 出版
北京东黄城根北街 16 号
邮政编码: 100717
http://www.sciencep.com

丽 源 印 刷 厂 印刷

科学出版社发行 各地新华书店经销
*
2008 年 6 月第 一 版 开本: B5(720 × 1000)
2008 年 6 月第一次印刷 印张: 37 3/4
印数: 1—3 000 字数: 709 000

定价: 88.00 元

(如有印装质量问题，我社负责调换〈环伟〉)

中文版序

本书的素材大致可以分为三部分. 第一部分给出所需的凸分析工具, 它们构成现代优化理论的基础. 具体地, 我们详尽地讨论了共轭对偶性、多值函数理论与微分学基础中的某些方面. 在第二部分, 这些工具被用于发展一阶与二阶最优性条件与参数最优化问题的最优值与最优解的扰动/变分理论. 本书最后几章处理这些一般性的结果在非线性规划、半定规划、半无限规划与最优控制中的应用. 对这些领域感兴趣的读者可以先阅读这些章节的内容.

尽管最近几年在随机规划与最优控制理论领域取得了有价值的进展, 我们相信本书仍然可以作为对优化理论感兴趣的学生和研究者一本有用的入门书.

我们期望本书的中文译本对这些素材在新一代学生中的传播起到帮助作用, 并带来最优化理论与应用的新的令人振奋的发展. 大连理工大学应用数学系张立卫教授在翻译过程中承担了大量的工作, 我们对此表示感谢, 同时感谢新加坡国立大学孙德锋教授的支持与鼓励.

J.F. 博南
INRIA and Ecole Polytechnique, France
A. 夏皮罗
Georgia Institute of Technology, USA
2007 年 8 月 20 日

中译本序

《最优化问题的扰动分析》一书由两位国际著名的优化专家Bonnans和Shapiro于2000年出版. 该书系统地介绍了优化和变分问题理论方面的最新成果. 其中, 对于非线性锥优化的阐述尤其深刻和全面. 该书是通向现代优化的入门工具, 对广大年轻的优化工作者具有十分重要的指导意义.

该书的中文版是由大连理工大学的张立卫教授历经数年几易其稿翻译而成. 该书的顺利出版必将帮助国内的优化研究迅速达到国际前沿并做出一系列领先的成果.

孙德锋
新加坡国立大学数学系
2007年10月20日

译 者 序

本书的两位作者 Bonnans 教授和 Shapiro 教授对译者的翻译工作非常支持, 他们把最新的勘误表寄给译者并为中文译本欣然作序. 新加坡国立大学的孙德锋博士一直关注着本书的翻译和出版工作, 给译者持久的鼓励和支持, 并在百忙之中抽出时间为中文译本写序, 希望中文译本的出版对我国最优化的发展起到帮助作用. 我对以上三位专家表示感谢.

我的博士研究生任咏红、于洪霞、金丽、肖现涛、顾剑、李阳、王韵, 硕士研究生李洁、刘溪、王百青、于桂花等用 CTEX 录入中文译本的原稿, 对他们表示感谢.

由于译者水平所限, 译文的某些地方难免会有不妥之处, 欢迎广大读者批评和指正.

本书中译本的出版得到了国家自然科学基金 (基金项目编号 10771026) 的资助, 在此表示感谢. 最后, 尤其要感谢科学出版社对本书出版所给予的支持.

张立卫
大连理工大学应用数学系
2007 年 10 月 25 日

基 本 记 号

基本集合与空间

$:=$　定义符号

$\equiv$　恒等于

$\varnothing$　空集

$|I|$　集合 I 的元素个数

$x \mapsto f(x)$　点 x 到 $f(x)$ 的映射

$\overline{\mathbb{R}}$　$= \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$ 增广实线

$\mathbb{R}^n$　n 维欧氏空间

$\mathbb{R}^n_+$　$= \{x \in \mathbb{R}^n : x_i \geqslant 0,\ i = 1, \cdots, n\}$ 非负卦限

$\mathbb{R}^n_-$　$= -\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x_i \leqslant 0,\ i = 1, \cdots, n\}$ 非正卦限

X, Y　Banach 或局部凸的拓扑向量空间

$\mathcal{S}^p$　$p \times p$ 对称矩阵构成的线性空间

$\mathcal{S}^p_+(\mathcal{S}^p_-)$　$p \times p$ 的正 (负) 半定矩阵构成的锥

$\mathcal{W}_r$　$\subset \mathcal{S}^p$ 秩为 r 的矩阵的集合

ℓ_2　满足 $\sum_{i=1}^{\infty} x_i^2 < \infty$, $\|x\| = \left(\sum_{i=1}^{\infty} x_i^2\right)^{1/2}$ 与 $\langle x, y\rangle = \sum_{i=1}^{\infty} x_i y_i$, $x, y \in \ell_2$ 的序列 $x = (x_i)_{i=1}^{\infty}$ 构成的 Hilbert 空间

$L_2[0,1]$　平方可积的实值函数 $\psi(t)$ 的等价类构成的 Hilbert 空间, 其中 $\psi_1 \sim \psi_2$ 的含义是对除了可能的零 Lebesgue 测度集合外的所有 $t \in [0,1]$, 均有 $\psi_1(t) = \psi_2(t)$, 且$\langle \psi, \phi\rangle = \int_0^1 \psi(t)\phi(t)\mathrm{d}t$

$L_P(\Omega, \mathcal{F}, \mu)$　具有有限范数 $\|\psi\|_p := \left(\int_\Omega |\psi(\omega)|^p \mathrm{d}\mu(\omega)\right)^{1/p}$ 的 $\mathcal{F}$ 可测函数 $\psi : \Omega \to \mathbb{R}$ 构成的 Banach 空间

$[L_p(\Omega)]_+$　$\subset L_p(\Omega, \mathcal{F}, \mu)$ 是几乎处处非负值函数的集合

$C(\Omega)$　定义于紧致度量空间 Ω 的赋予上确界范数 $\|\psi\| = \sup_{\omega \in \Omega} |\psi(\omega)|$ 的连续函数 $\psi : \Omega \to \mathbb{R}$ 构成的 Banach 空间

$C^l(\Omega)$　l 次连续可微函数 $\psi : \Omega \to \mathbb{R}$ 的 Banach 空间, 其中 $\Omega \subset \mathbb{R}^n$

$C^{1,1}(\Omega)$　$D\psi(\cdot)$ 是局部 Lipschitz 连续的可微函数 $\psi : \Omega \to \mathbb{R}$ 的空间

$\mathcal{D}(\Omega)$　在 Ω 上定义的具有紧致支撑的实值的 C^∞ 光滑函数的集合

$C_{00}(\Omega)$　在 Ω 中具有紧致支撑的连续函数的集合

$\mathcal{O}_{K,m}$　与在 $\mathcal{D}(\Omega)$ 上的拓扑相联系的桶集族

$\mathcal{O}_K$	与在 $C_{00}(\Omega)$ 上的拓扑相联系的桶集族
$W^{m,s}(\Omega)$	$=\{\psi\in L_s(\Omega):D^q\psi\in L_s(\Omega)$,若$\|q\|\leqslant m\}$Sobolev 空间, 其中 $D^q\psi=\partial^{\|q\|}\psi/\partial x_1^{q_1}\cdots\partial x_l^{q_l}$ 且 $\|q\|=q_1+\cdots+q_l$
$W_0^{m,s}(\Omega)$	在 $W^{m,s}(\Omega)$ 中 $\mathcal{D}(\Omega)$ 的闭包
$W^{1,\infty}(\Omega)$	Lipschitz 连续函数 $\psi:\Omega\to\mathbb{R}$ 的 Banach 空间
$H^m(\Omega)$	Sobolev 空间$W^{m,2}(\Omega)$
$H^{-1}(\Omega)$	$H_0^1(\Omega)$ 的对偶空间
$W^{-1,s'}(\Omega)$	$W_0^{1,s}(\Omega)$ 的对偶空间
$C_+(\Omega)$	空间 $C(\Omega)$ 中的非负值函数的集合
$C_-(\Omega)$	空间 $C(\Omega)$ 中的非正值函数的集合
$\mathcal{L}(X,Y)$	赋予算子范数 $\\|A\\|=\sup\limits_{x\in B_X}\\|Ax\\|$ 的线性的连续算子 $A:X\to Y$ 构成的空间
X^*	$=\mathcal{L}(X,\mathbb{R})$, X 的对偶空间
$B(x,r)$	$=\{x'\in X:\\|x'-x\\|<r\}$, 中心为 x 的半径为 $r>0$ 的开球
B_X	$=B_X(0,1)$, X 的单位开球
$\overline{B}_X$	X 的闭单位球
$[\\![x]\\!]$	$=\{tx:t\in\mathbb{R}\}$,由 x 生成的线性空间
2^X	X 的子集构成的集合
$\dim(X)$	线性空间 X 的维数
$\mathcal{P}_\Omega$	$=\{\mu\in C(\Omega)^*:\mu(\Omega)=1,\mu\succeq 0\}$, Ω上的概率测度的集合
$\mathrm{cap}(A)$	集合 A 的容量 (capacity)

矩阵与向量

$\langle\alpha,x\rangle$	线性泛函 $\alpha\in X^*$ 在 $x\in X$ 处的值
$x\cdot y$	$=\sum\limits_{i=1}^n x_iy_i$ 两个有限维向量 $x,y\in\mathbb{R}^n$ 的数积
A^{T}	矩阵 A 的转置
$\mathrm{rank}(A)$	矩阵 A 的秩
$\mathrm{vec}(A)$	矩阵 A 的列拉直得到的向量
$A^\dagger$	矩阵 A 的 Moore-Penrose 伪逆
$\mathrm{trace}A$	$=\sum\limits_{i=1}^p a_{ii}$, $p\times p$ 矩阵 A 的迹
$A\cdot B$	$=\mathrm{Tr}(AB)$, 两个对称矩阵 $A,B\in\mathcal{S}^p$ 的数积
$A\otimes B$	A 与 B 的 Kronecker 积
$\lambda_{\max}(A)$	对称矩阵 $A\in\mathcal{S}^p$ 的最大特征值
$A\succeq 0(A\preceq 0)$	矩阵 $A\in\mathcal{S}^p$ 是正 (负) 半定的
I_p	$p\times p$ 单位矩阵

集合上的运算

$\mathrm{Sp}(S)$	$=\mathbb{R}_+(S-S)$ 由集合 $S\subset X$ 生成的线性空间
$\mathbb{R}_+(S)$	$=\{tx: x\in S, t\geqslant 0\}$ 集合 $S\subset X$ 生成的锥
$\mathrm{cl}(S)$	集合 $S\subset X$ 的拓扑包, 若 X 是 Banach 空间, 闭包是关于范数 (即强) 拓扑取的
$\mathrm{int}(S)$	$=\{x\in S: \text{存在}x\text{的邻域}V\text{满足}V\subset S\}$, 集合 S 的内部
$\mathrm{bdr}(S)$	(也记为 ∂S)$=\mathrm{cl}(S)\setminus\mathrm{int}(S)$, 集 S 的边界
$\mathrm{ri}(S)$	$=\{x\in S: \text{存在 } x \text{ 的邻域 } V \text{ 满足 } V\cap(x+L)\subset S\}$(其中 $L:=\mathrm{cl}\,\mathrm{Sp}(S)$) 凸集 S 的相对内部
$\mathrm{core}(S)$	$=\{x\in S: \forall x'\in X, \exists\varepsilon>0, \forall t\in[-\varepsilon,\varepsilon], x+tx'\in S\}$
$\mathrm{dist}(x,S)$	$=\inf\limits_{z\in S}\\|x-z\\|$, $x\in X$到集合 $S\subset X$ 的距离
$\mathrm{Haus}(S,T)$	$=\max\left\{\sup\limits_{x\in S}\mathrm{dist}(x,T), \sup\limits_{x\in T}\mathrm{dist}(x,S)\right\}$, 集合 S 与 T 间的 Hausdorff 距离
$S^{\perp}$	$=\{\alpha\in X^*: \langle\alpha,x\rangle=0, \forall x\in S\}$, 集合 $S\subset X$ 的直交补
S^{∞}	$=\{h\in X: \exists x\in S, \forall t\geqslant 0, x+th\in S\}$, 凸集 S 的回收锥
$\sigma(\alpha,S)$	$=\sup\limits_{x\in S}\langle\alpha,x\rangle$, 集合 S 的支撑函数
$I_S(\cdot)$	集合 S 的指示函数
$\mathrm{conv}(S)$	集合 S 的凸包
$\mathrm{diam}(S)$	$=\sup\limits_{x,x'\in S}\\|x-x'\\|$, 集合 S 的直径
C^-	$=\{\alpha\in X^*: \langle\alpha,x\rangle\leqslant 0, \forall x\in C\}$, 锥 $C\subset X$ 的极 (负对偶), 其中 X 与 X^* 是成对的空间
$\mathrm{lin}(C)$	凸锥 C 的线子空间
$a\preceq_C b$	由锥 C 引导的序关系, 即 $b-a\in C$
$a\vee b$	a 与 b 的最小上界
$a\wedge b$	a 与 b 的最大下界
$[a,b]_C$	$=\{x: a\preceq_C x\preceq_C b\}$, 相对于序关系 "$\preceq_C$" 的区间
$G\,\overline{\pitchfork}_x\,W$	意味着映射 G 在点 x 处与流形 W 横截的相交

切　　集

$T_S(x)$	$=\limsup\limits_{t\downarrow 0}\dfrac{S-x}{t}$, 集合 S 在点 $x\in S$ 处的余 (bouligand) 切锥
$T_S^i(x)$	$=\liminf\limits_{t\downarrow 0}\dfrac{S-x}{t}=\{h\in X: \mathrm{dist}(x+th,S)=o(t), t\geqslant 0\}$, 集合 S 在点 $x\in S$ 处的内切锥
$T_S^c(x)$	集合 S 在 $x\in S$ 处的 Clarke 切锥
$\mathcal{R}_S(x)$	$=\{h\in X: \exists t>0, x+th\in S\}$, 凸集 S 在点 $x\in S$ 的雷达锥
$T_S(x)$	$=\mathrm{cl}[\mathcal{R}_S(x)]=T_S^i(x)$, 凸集 S 在点 $x\in S$ 处的切锥

$T_S^2(x,h)$	$=\limsup\limits_{t\downarrow 0}\dfrac{S-x-th}{\frac{1}{2}t^2}$, 集合 S 在点 $x\in S$ 处沿方向 h 的外二阶切集
$T_S^{i,2}(x,h)$	$=\liminf\limits_{t\downarrow 0}\dfrac{S-x-th}{\frac{1}{2}t^2}$, 集合 S 在点 $x\in S$ 处沿方向 h 的内二阶切集
$T_S^{i,2,\sigma}(x,h)$	$=\lim\limits_{n\to\infty}\inf\dfrac{S-x-t_nh}{\frac{1}{2}t_n^2}$, 与序列 $\sigma=\{t_n\}$, $t_n\downarrow 0$ 相联系的序列二阶切集
Σ	收敛于零的正数序列 $\sigma=\{t_n\}$ 的集合
$N_S(x)$	$=[T_S(x)]^-$集合 $S\subset X$, 在 $x\in S$ 处的法锥
$N_S(x)$	$=\{\alpha\in X^*:\langle\alpha,z-x\rangle\leqslant 0,\forall z\in S\}$, 凸集 S 的法锥
$PN_S(x)$	S 在 x 处的追近法向量的集合
$PN_S^\delta(x)$	S 在 x 处的 δ 追近法向量的集合

函数与算子

f	$X\to\overline{\mathbb{R}}$, 增广实值函数
$\mathrm{dom}f$	$=\{x\in X:f(x)<+\infty\}$, 函数 f 的定义域
$\mathrm{gph}f$	$=\{(x,f(x)):x\in X\}\subset X\times\mathbb{R}$, 函数 f 的图
$\mathrm{epi}f$	$=\{(x,\alpha):\alpha\geqslant f(x),x\in X\}\subset X\times\mathbb{R}$, 函数 f 的上图
$\mathrm{lsc}f(x)$	$=\min\{f(x),\liminf\limits_{x'\to x}f(x')\}$, f 的下半连续包
$\mathrm{cl}f(x)$	函数 f 的闭包
$\mathrm{conv}f$	函数 f 的凸包
$\mathrm{lev}_{\leqslant\alpha}f$	$=\{x\in X:f(x)\leqslant\alpha\}$, 函数 f 的水平集
$f^*(x^*)$	$=\sup\limits_{x\in X}\{\langle x^*,x\rangle-f(x)\}$, 函数 f 的共轭
$\hat{f}_\varepsilon(\cdot)$	函数 f 的 Moreau-Yosida 正则化
$f\diamond g(u)$	$=\inf\limits_{x\in X}\{f(u-x)+g(x)\}$, 增广实值函数 $[f,g:X\to\overline{\mathbb{R}}]$ 的下卷积
$f\circ g$	映射 $g:X\to Y$ 与映射 (增广实值函数)$f:Y\to Z$ 的复合, 即 $(f\circ g)(x)=f(g(x))$
$\lambda^\perp$	$=\ker\lambda=\{y\in Y:\langle\lambda,y\rangle=0\}$, $\lambda\in Y^*$ 的零空间
$\mathcal{N}(Q)$	$=\{x\in X:Q(x)=0\}$, 二次形式 $Q:X\to\mathbb{R}$ 的零空间
$A^*:$	$Y^*\to X^*$, 连续的线性算子 $A:X\to Y$ 的伴随算子, 即对所有的 $x\in X$ 与 $\lambda\in Y^*$, $\langle A^*\lambda,x\rangle=\langle\lambda,Ax\rangle$
$\prod_S(x)$	$=\mathrm{argmin}\limits_{z\in S}\lVert x-z\rVert$, 点 x 到 S 上的集值度量投影
$P_S(x)$	$\in\prod_S(x)$, 点 x 到 S 上的度量投影
$\triangle y$	$=\sum\limits_{i=1}^n\dfrac{\partial^2y}{\partial\omega_i^2}$, Laplace 算子
$\delta(\omega)$	在点 ω 处的质量 1 的测度 (Dirac 测度)

$\mu \geqslant 0$	意味测度 μ 是非负值的
$\mathrm{supp}(\mu)$	测度 μ 的支撑
$\|\mu\|$	测度 μ 的全变差
$[a]_+$	$=\max\{0,a\}, a\in\mathbb{R}$
$\forall$	对任意的
$\exists$	存在

多 值 函 数

$\Psi:$	$X\to 2^Y$, 多值函数 (点到集合映射), 它将 X 映到 Y 的子集构成的集合上
$\mathrm{dom}\Psi$	$=\{x\in X:\Psi(x)\neq\varnothing\}$, Ψ 的定义域
$\mathrm{range}(\Psi)$	$=\Psi(X)=\{y\in Y:y\in\Psi(x),x\in X\}$, Ψ 的值域
$\mathrm{gph}(\Psi)$	$=\{(x,y)\in X\times Y:y\in\Psi(x),x\in X\}$, Ψ 的图
$\Psi^{-1}(y)$	$=\{x\in X:y\in\Psi(x)\}$, Ψ 的逆多值函数
$\lim\limits_{x\to x_0}\sup\Psi(x)$	$=\{y\in Y:\liminf\limits_{x\to x_0}[\mathrm{dist}(y,\Psi(x))]=0\}$, 多值函数 Ψ 在 x_0 处的上集合极限
$\lim\limits_{x\to x_0}\inf\Psi(x)$	$=\{y\in Y:\limsup\limits_{x\to x_0}[\mathrm{dist}(y,\Psi(x))]=0\}$, 多值函数 Ψ 在 x_0 处的下集合极限

极限与导数

$r(h)=o(h)$	意味着当 $h\to 0$ 时 $r(h)/\|h\|\to 0$
$r(h)=O(h)$	意味着对 $0\in X$ 的邻域中的所有 h, $r(h)/\|h\|$ 是有界的
$\nabla f(x)$	$=\left(\dfrac{\partial f(x)}{\partial x_1},\cdots,\dfrac{\partial f(x)}{\partial x_n}\right)$, 函数 $f:\mathbb{R}^n\to\mathbb{R}$ 在 $x\in\mathbb{R}^n$ 处的梯度
$\nabla^2 f(x)$	$=\left[\dfrac{\partial^2 f(x)}{\partial x_i\partial x_j}\right]_{i,j=1}^n$, 函数 $f:\mathbb{R}^n\to\mathbb{R}$ 在 $x\in\mathbb{R}^n$ 处的二阶偏导数构成的 Hesse 阵
$Dg(x):$	$X\to Y$, 映射 $g:X\to Y$ 在点 $x\in X$ 处的 (Gâteaux,Hadamard 或 Fréchet, 依赖于上下文) 导数
$D^2g(x):$	$X\to\mathcal{L}(X,Y)$, 映射 g 在点 x 处的二阶导数
$D^2g(x)(h,h)$	$=[D^2g(x)h]h$, 对应于 $D^2g(x)$ 的二次型
$D_xg(x,u)$	映射 $g:X\times\mathcal{U}\to Y$ 的偏导数
$g'(x,d)$	$=\lim\limits_{t\downarrow 0}\dfrac{g(x+td)-g(x)}{t}$, 映射 $g:X\to Y$ 在点 x 处沿方向 d 的方向导数
$f'_+(x,d)$	$=\limsup\limits_{t\downarrow 0}\dfrac{f(x+td)-f(x)}{t}$, 函数 $f:X\to\overline{\mathbb{R}}$ 的上方向导数
$f'_-(x,d)$	$=\liminf\limits_{t\downarrow 0}\dfrac{f(x+td)-f(x)}{t}$, 函数 $f:X\to\overline{\mathbb{R}}$ 的下方向导数

$f''(x;d,w)$ $\quad = \lim\limits_{t\downarrow 0} \dfrac{f\left(x+td+\frac{1}{2}t^2w\right)-f(x)-tf'(x,d)}{\frac{1}{2}t^2}$，函数 f 的二阶方向导数

$f_-^{\downarrow}(x,h)$ $\quad = \mathrm{e}-\liminf\limits_{t\downarrow 0} \dfrac{f(x+th)-f(x)}{t}$，下方向上图导数

$f_+^{\downarrow}(x,h)$ $\quad = \mathrm{e}-\limsup\limits_{t\downarrow 0} \dfrac{f(x+th)-f(x)}{t}$，上方向上图导数

$f_-^{\downarrow\downarrow}(x;h,w)$ $\quad = \mathrm{e}-\liminf\limits_{t\downarrow 0} \dfrac{f\left(x+th+\frac{1}{2}t^2w\right)-f(x)-tf_-^{\downarrow}(x,h)}{\frac{1}{2}t^2}$，下二阶方向上图导数

$f_+^{\downarrow\downarrow}(x;h,w)$ $\quad = \mathrm{e}-\limsup\limits_{t\downarrow 0} \dfrac{f\left(x+th+\frac{1}{2}t^2w\right)-f(x)-tf_+^{\downarrow}(x,h)}{\frac{1}{2}t^2}$，上二阶方向上图导数

$d^2f(x|\alpha)(h)$ $\quad := \liminf\limits_{\substack{t\downarrow 0\\ h'\to h}} \dfrac{f(x+th')-f(x)-t\langle\alpha,h'\rangle}{\frac{1}{2}t^2}$，函数 f 在点 x 处关于 $\alpha\in X^*$ 的二阶次导数

$\partial f(x)$ $\quad = \{x^*\in X^*: f(y)-f(x)\geqslant\langle x^*,y\rangle,\forall y\in X\}$，函数 $f: X\to\overline{\mathbb{R}}$ 的次微分

最优化问题

val(P) 问题 (P) 的最优值

Φ 问题 (P) 的可行集

$\mathcal{S}(\mathrm{P})$ 问题 (P) 的最优解集

$L(x,\lambda)$ $= f(x)+\langle\lambda,G(x)\rangle$，问题 (P) 的 Lagrange 函数

$L^g(x,\lambda)$ $= \alpha f(x)+\langle\lambda,G(x)\rangle$，广义的 Lagrange 函数

$L^s(x,\lambda)$ $= \langle\lambda,G(x)\rangle$，奇异的 Lagrange 函数

$\Lambda(x)$ x 处的 Lagrange 乘子的集合

$\Lambda^g(x)$ x 处的广义 Lagrange 乘子的集合

$\Lambda^s(x)$ x 处的奇异 Lagrange 乘子的集合

$\Lambda_N^g(x)$ $= \{(\alpha,\lambda)\in\Lambda^g(x):\alpha+\|\lambda\|=1\}$，$x$ 处的正规化的广义 Lagrange 乘子的集合

$I(x)$ $= \{i: g_i(x)=0, i=q+1,\cdots,p\}$，在 x 处的起作用的不等式约束的集合

$I_+(x,\lambda)$ $= \{i\in I(x):\lambda_i>0\}$

$I_0(x,\lambda)$ $= \{i\in I(x):\lambda_i=0\}$

$\triangle(x)$ $= \{\omega\in\Omega: g(x,\omega)=0\}$，$x$ 处的起作用的约束 $g(x,\omega)\leqslant 0,\ \omega\in\Omega$ 的集合

$C(x)$ 在点 x 处临界方向的集合 (临界锥)

$C_\eta(x)$	在点 x 处的近似临界锥
(P_u)	以 $u \in U$ 为参数的最优化问题
$\Phi(u)$	参数化问题 (P_u) 的可行集
$v(u)$	$= \mathrm{val}(\mathrm{P}_u) = \inf\limits_{x\in\Phi(u)} f(x,u)$, (P_u) 的最优值 (边际 (marginal)) 函数
$\mathcal{S}(u)$	$= \mathcal{S}(\mathrm{P}_u) = \operatorname*{argmin}\limits_{x\in\Phi(u)} f(x,u)$, (P_u) 的最优解集
$\bar{x}(u)$	$\in \mathcal{S}(u)$ (P_u) 的最优 (ε 最优) 解
$L(x,\lambda,u)$	(P_u) 的 Lagrange 函数
$L^g(x,\lambda,u)$	(P_u) 的广义 Lagrange 函数
$L^s(x,\lambda,u)$	(P_u) 的奇异 Lagrange 函数
$\Lambda(x,u)$	问题 (P_u) 的 Lagrange 乘子的集合
$\Lambda^g(x,u)$	问题 (P_u) 的广义 Lagrange 乘子的集合

目　录

第1章 引　　言

本书主要研究连续最优化问题的扰动分析. 在最近二十年里, 这一领域取得了重要进展, 现在应该是对应用于各类问题得到的众多重要结果给出综合评述的时候了. 本书所考虑的模型问题具有下述形式

$$\text{(P)} \qquad \min_{x\in X} f(x) \quad \text{s.t.} \quad G(x) \in K,$$

这里 X 与 Y 是 Banach 空间, K 是 Y 中的闭凸子集, $f: X \to \mathbb{R}$, $G: X \to Y$ 分别是目标函数与约束映射. 我们也考虑上述问题的参数化形式 (P_u), 其中目标函数 $f(x,u)$ 与约束映射 $G(x,u)$ 中的向量 u 为参数, u 在 Banach 空间 U 中变化. 我们的目的是, 在将 (P_u) 的最优值函数 $v(u)$ 与最优解集合 $\mathcal{S}(u)$ 视为参数向量 u 的函数与映射时, 研究它们的连续性质与可微性质.

几类重要的最优化问题可以模型化为上述较抽象的框架中.

• 非线性规划. 即 (P) 是有限维的最优化问题, 它的可行集由有限多个等式与不等式类型的约束来定义. 上述抽象的形式适合于表达这样的问题, 其中 $K := \{0_q\} \times \mathbb{R}_-^{p-q}$, 0_q 是空间 $\mathbb{R}^q$ 的零向量, $\mathbb{R}_-^{p-q}$ 是 $\mathbb{R}^{p-q}$ 的负象限.

• 半定规划. 这也是有限维的最优化问题, 其约束为对称矩阵值映射 $G(x)$ 是半正定的, 即空间 X 是有限维的, 空间 Y 定义为 $p \times p$ 对称矩阵构成的线性空间, 集合 K 是由 $p \times p$ 半正定矩阵构成的锥. 在许多应用中, 目标函数 $f(x)$ 是线性的, 约束映射 $G(x)$ 是仿射的, 这样的问题有时被称为线性矩阵不等式 (LMI).

• 半无限规划. 这又是一个有限维最优化问题的例子, 其中空间 X 是有限维的, 具有无穷多个不等式约束 $g(x,\omega) \leqslant 0$, 其中 ω 是紧致集合 Ω 中的参数. 若 $g(x,\omega)$ 是连续的, 则通过定义 Y 为 Ω 上的连续函数的空间, $G(x)(\cdot) = g(x,\cdot)$, 我们可以在上述抽象框架下考虑这样的问题.

• 最优控制. 一般说来, 这涉及到由微分方程支配的 (governed) 系统, 通过某些控制变量的最优化, 它的解被称为状态的. 本书不去处理最优控制理论的某些具体专题, 如 Pontryagin 原理或 Hamilton-Jacobi-Bellman 方程, 这些问题的微分或动态结构已有深入研究. 另一方面, 我们证明, 本书所发展的理论可以推导二阶必要性/充分性的最优性条件且可以进行灵敏度分析. 我们通过考虑被半线性椭圆方程支配的系统的最优控制这个例子来讨论这些问题.

• 力学中的变分问题. 许多静态力学中的接触问题可视为在物体变形的约束之下某一势能的极小化问题. 我们详细分析作为障碍问题的简单模型问题.

• 复合与极小–极大最优化. 复合参数化问题是下述形式的最优化问题

$$\min_{x\in X} F(G(x,u)),$$

其中 $F: Y\to\overline{\mathbb{R}}$ 是下半连续的增广实值的凸函数, $G: X\times U\to Y$ 是映射. 这一问题可表示为所考虑的问题 (P_u) 的形式. 相反的结论也是成立的, 因此这两类问题是等价的. 所以, 选择哪一个问题作为研究对象只是一个爱好或便利的问题.

若 $Y:=C(\Omega)$, 函数 F 是极大值的形式, $F(y):=\sup\limits_{\omega\in\Omega} y(\omega)$, $G(x,u)(\cdot)=g(x,u,\cdot)$, 上述复合问题变为极小–极大问题

$$\min_{x\in X}\max_{\omega\in\Omega}\quad g(x,u,\omega).$$

相关领域

另外, 还有一些其他类的问题, 尽管不是问题 (P) 框架下的具体情况, 它们也是密切相关的.

• 变分不等式与广义方程. 考虑下述问题：求 $x\in K$ 满足

$$\text{(VI)}\qquad \langle F(x), x'-x\rangle\geqslant 0,\quad \forall x'\in K.$$

若 $F(\cdot)$ 是实值函数 $f(\cdot)$ 的导数, 则 (VI) 表示受约束于 $x\in K$ 的极小化 $f(x)$ 问题的一阶最优性条件. 然而, 在力学 (非均质材料) 与经济学中, 存在不能简化为最优化问题的各种各样的例子. 我们给出的某些方法适合于变分不等式 (VI) 的扰动分析, 见 5.1 节.

• 最优化算法. 在研究求解最优化问题的数值算法的收敛性质时, 二阶分析是重要的. 所以, 尽管没有显式地处理这样的专题, 本书对这样的分析也给出了有用的素材.

光滑性与非光滑性

本书的绝大部分均假设目标函数 f 与约束映射 G 均是光滑的. 非光滑函数的微分学的最新进展令人印象深刻, 读者会好奇我们为什么选择 f 与 G 的光滑性质. 对此有如下的动因：

• 一个技术性原因是计算最优化问题最优解的导数需要二阶分析. 因此, 放宽 f 与 G 的光滑性假设条件会使分析大大复杂化.

• 几类很类似的问题可以很自然地表示为这样的框架之下.

• 这使我们可集中到本质性的困难上, 即集合 K 的二阶分析, 在某种意义下, 即集中到问题的非光滑性上.

应该注意到, 集合 K 可以是任意的闭凸集合. 因此, 某些非光滑问题也可以放在这样一个框架之下进行研究. 这样的一大类问题即复合最优化问题类, 如前所述, 其中函数 $F(\cdot)$ 被设为是凸函数, 但不必是光滑的. 进一步, 这里发展起来的某些技术也可以以直接的方式用于研究非光滑问题 (如 4.10.3 节中关于泛函空间的二阶展开).

这样的研究视点被证明是颇具成果的. 它还可以使我们对未解的问题有一个好的观察视角. 这里给出的结果推广到非光滑的情形是很有意义的问题, 可能是将来研究的一个主题.

二阶最优性条件

一阶尤其二阶最优性条件与灵敏度分析有着较强的联系. 事实上, 作为灵敏度分析研究的副产品, 可得到各种各样的关于最优性条件的结果. 还有, 这两种理论中未解决的问题也具有较强的相似性.

除了假设集合 K 是闭凸集外, 在没有其他具体假设的情况下, (在某些约束规范之下) 可以叙述一些二阶必要性或二阶充分性条件. 然而, 通常而言, 必要性条件与充分性条件之间存在差别, 有两个原因, 其一是集合 K 的二阶性质 (曲率), 其二是对于无穷维问题缺乏某些紧致性. 后一种情况出现在假设 Lagrange 函数的 Hesse 阵是 Legendre 形式的许多应用中, 见定义 3.73. 有几类令人感兴趣的问题, 可以得到与二阶充分的最优性条件间 “无间隙” 的二阶必要性条件.

称在可行点 x_0 处的二阶最优性条件的理论是完备的, 若可用数据的二阶信息来刻画下述的二阶增长条件: 存在 $c > 0$ 满足对充分接近于 x_0 的所有的可行点 x, 成立

$$f(x) \geqslant f(x_0) + c\|x - x_0\|^2. \tag{1.1}$$

存在几类问题, 其二阶最优性条件的理论是完备的, 如非线性规划、半定规划、半无限规划(在某些正则性条件之下)、具有二阶正则性约束的问题 (见 3.3.3 节)、3.2.3 节意义下约束的多面性质的问题.

可以得到在可行点 x_0 的某一邻域中局部最优解的稳定性与上述二阶增长条件的一致形式是密切相关的 (见 5.1.3 节). 仅仅对几种具体情况, 这样的一致二阶增长性的一组完备的 (即必要且充分的) 二阶最优性条件是已知的, 这样的条件可能是进一步研究的一个专题.

稳定性分析

本书的一个核心专题是参数化问题 (P_u) 的灵敏度分析. 由于最优值函数和最优解 $\bar{x}(u) \in \mathcal{S}(\mathrm{P}_u)$ 是典型的不可微的, 我们集中关注参数向量 u 沿路径 $u(t) := u_0 + td + o(t)$ 的方向分析, 其中 $d \in U$ 是给定的方向. 为了进行这样的方向分析,

需要约束规范 (或至少为约束规范的方向形式, 见定义 4.8) 与某些二阶充分性条件, 使得目标函数的充分增长性成立. 如果忽略集合 K 可能的曲率, 在某些情况下, 最优值函数的二阶上估计与二阶下估计间存在间隙. 这一间隙同相对应的二阶必要性与二阶充分性间的间隙具有相类似的本质.

理论的关键之处在于要填充这一间隙. 有两大类问题, 对它们而言, 所给出的理论可成功地填充这一间隙. 其中一种情况是空间为 X, 是有限维的, 集合 K 具有某一性质, 这一性质是与二阶切集相关的, 称为二阶正则性(二阶正则性还用于推导无间隙二阶最优性条件). 在某一程度下, 这一概念使得我们可以对非线性规划问题、半定规划问题、对半无限规划问题给出一个完整的灵敏度分析. 当 X 是无限维空间时, 情况变得不同. 然而, 在 Legendre 形式与广义的多面性质的框架之下, 可以推导最优值函数的二阶展开及最优 (或接近最优) 解的相应的展开.

扰动性分析的主要结果可以区分 5 种情况. 其中两种情况的假设是相当强的. 第一种情况是分析可简化为将隐函数定理应用到一阶最优性系统的情形. 这是相对简单的情况, 适合于等式问题, 也适合有有限个不等式约束且严格互补性假设成立的情况. 这一类型的结果在文献 [72] 中有较深入的讨论, 其理论可以推广用于处理半定规划问题, 见 5.3.6 节. 第二种情况是一阶最优性系统满足由 Robinson[175] 提出的强正则性假设的情形. 粗略地说, 这意味着扰动最优性系统具有以 Lipschitz 方式依赖于扰动的局部的唯一解. 此种情况, 最优化问题的灵敏度分析简化为最优性系统的灵敏度分析. 强正则性的方法可视为隐函数定理方法的一个推广. 对于非线性规划而言, 这尤其有意思, 因为在这种情况下, 可以知道如何用数据的一阶与二阶展开来刻画强正则性.

本书给出的其他三个基本情况的分析具有不同的本质, 基于相当弱的假设. 分析的关键在于在方向正则性与二阶充分性条件的某一形式之下, 计算最优值函数的上方与下方的二阶方向估计. 这一方法需应用凸分析与对偶理论的强有力工具.

这三种情形的第一种是强方向二阶充分性条件成立的情况. 此种情况, 可以计算最优值函数的二阶展开与接近最优解的一阶展开. 第二种就是标准的二阶充分性条件成立的情形. 可以计算最优值函数的一阶展式与一个解路径的参数 t 开平方阶的展开. 第三种是在假设涉及奇异的 Lagrange 乘子二阶充分性条件成立的前提下, 处理 Lagrange 乘子集合为空集的情况. 可以计算第一项为参数 t 的开平方阶的最优值函数展开与解路径也是参数 t 的开平方阶的展开式.

具体应用

除上述提到的几类最优化问题, 我们还讨论下述的具体应用.

• 静态力学. 给出静态力学中的某些问题, 通过它们对非线性规划的扰动理论给出好的诠释. 我们详尽讨论了悬链线问题的两个奇异情形, 即悬链线是水平或者

竖直的两种情形.

• 电子网络的分析. 考虑直流电网络, 在网络的节点上输入电量是有限的约束之下, 讨论网络上的电量损失极小化的问题. 在非常高电压 (very high voltage) 的假设之下, 对问题进行扰动分析, 使得我们能够计算最优值函数与最优解的展开. 这一结果可以推广到最优电力流量问题, 除了将直流电替换为交流电外, 结论是类似的.

• 度量(正交) 投影. 设 S 是 Banach 空间 X 的闭子集. 对点 $x \in X$, 考虑 S 中与 x 最近的点的集合 $\Pi_S(x)$. 集合 $\Pi_S(x)$ 中选取的元素 $\mathrm{P}_S(x)$ 被称为 x 到 S 上的度量(直交) 投影. $\mathrm{P}_S(x)$ 的存在性、连续性与可微性是非常有意义的. 度量投影$\mathrm{P}_S(x)$ 可视为相对应的最优化问题的最优解. 当空间 X 是有限维的或 Hilbert 空间, 其连续性与可微性性质可放在我们的框架之下进行研究.

本书的框架

第 2 章处理泛函分析、微分分析、凸分析与对偶性理论的一般性结果. 给出的结果是非常基本的. 某些结果是经典的和众所周知的, 而其他的结果更具体些. 后面的扰动分析所基于的两个理论是共轭的 (参数) 对偶性理论与基于度量正则性概念的稳定性分析, 并将它们应用于广义线性规划. 这一章自成体系, 可以作为研究生课程 "非线性分析及其在最优化理论中的应用" 的基础.

第 3 章给出局部最优性的必要或 (且) 充分的条件. 这一章引入了后续的稳定性分析所用到的基本方法. 尤其讨论了下述专题: (可能是广义的)Lagrange 乘子的存在性、一阶充分性最优性条件、一般的二阶必要性条件及不含有集合 K 的曲率信息的 "标准的" 二阶充分性条件. 用二阶正则性与多面性质的概念说明如何消除这一间隙. 本章还包括复合最优化的讨论、非孤立最优解情况的分析以及某些具体专题, 如二次规划、简化过程与精确惩罚函数的分析. 本章给出一阶与二阶最优性条件的系统讨论, 对期望对这一专题有一般性认识的读者都是有吸引力的.

第 4 章给出最优化问题扰动理论的系统研究. 包含若干一般性结果, 包括如方向正则性与非退化性这样的一阶条件、最优值函数与相应的最优 (接近最优) 解集合与 Lagrange 乘子集合的量化的连续性性质. 还给出了最优值函数的一阶与二阶展开及最优解的一阶展开. 对三个 "基本情况" 的每一情况, 导出最优值函数展开的第一项的公式 (甚至在 "第一情况" 下得到第二项的公式), 在较强的条件下, 给出 (接近的) 最优解路径的估计, 使我们能够计算最优解路径的展开的第一项. 本章最后讨论连续函数空间中的一致近似与二阶分析.

第 5 章集中考虑具体的结构, 即变分不等式、非线性规划、半定规划与半无限规划. 感兴趣于其中某一个专题的读者可以先阅读本章. 非线性 (与线性) 规划的那一节自成体系, 仅用到有限维数空间的数学分析的标准工具, 关于半定规划与半无

限规划的其他两节, 尽管所给出的结果依赖于前面的进展, 但在某种程度上也是自成体系的.

第 6 章用于处理最优控制问题. 讨论偏微分方程最优控制的某些简单模型问题. 结果的一部分用抽象理论可以得到, 而其他结果则需要对前面章节给出的工具进行推广. 期望对最优控制问题的二阶理论有认识的读者可以先读此章.

同行 Maïtine Bergounioux, Alec Ioffe, Diethard Klatte, Boris Mordukhovich 与 Penot J-P 对本书给出了极有帮助的评述, 我们对此表示感谢. 对 Springer 出版社的良好的工作也表示感谢, 尤其是对本书编辑 Achi Dosanjh 表示谢意. 当然, 我们极大地获益于与合作者的工作, 尤其是与 Eduardo Casas, Roberto Cominetti, Alec Ioffe 以及 Housnaa Zidani 的合作工作.

第2章 背景素材

本章回顾拓扑学与泛函分析中的某些基本结果以及在凸和非凸优化问题的扰动理论中起实质性作用的某些数学工具. 我们处理的所有优化问题都是在 Banach 空间的框架之下进行的, 这里描述的某些结果则是在相当一般的局部凸拓扑向量空间的框架下进行的. 这样做的原因是 (赋予强拓扑的) Banach 空间 X 是 (赋予弱 * 拓扑的) 对偶空间 X^* 的对偶空间. 在局部凸空间的体系下, 原始空间与对偶空间存在着完整的对称结构. 因此, 两个方向上都可获得对偶性结果. 这里, 对某些结果给出证明, 而其他的 (经典) 结果只给出其描述. 它们的证明几乎在任何一本泛函分析的教科书中均可找到.

2.1 基本泛函分析

2.1.1 拓扑向量空间

1. *拓扑空间*

设 E 是集合. 记号 $e \in E$ 或 $E \ni e$ 表示 e 是 E 中的元素 (element), 若 F 中的每一元素均是 E 中的元素, 则称 F 是 E 的子集 (subset), 记为 $F \subset E$ 或 $E \supset F$. 集合 F 在 E 中的补集 (complement) 是 $E \setminus F := \{e \in E : e \notin F\}$. 用 2^E 记 E 中的所有子集的集合.

集合 E 上的一个拓扑(topology) 是满足下述三条公理的 E 的子集类 $\mathcal{G} \subset 2^E$:

(i) E 与 $\varnothing$(空集) 属于 $\mathcal{G}$.

(ii) $\mathcal{G}$ 中的任何元素的并属于 $\mathcal{G}$, 且

(iii) $\mathcal{G}$ 中任何有限个元素的交属于 $\mathcal{G}$.

$\mathcal{G}$ 中的元素被称为开集(open set), $(E, \mathcal{G})$ 被称为一个拓扑空间(topological space). 开集在 E 中的补集被称为闭集 (closed set). 注意到, E 与 $\varnothing$ 是闭集, 有限多个闭集的并及 (可能无穷多个) 闭集的交是闭集.

元素 $e \in E$ 的邻域(neighborhood) 是 E 的子集, 它包含含有 e 的开集. 设 F 是 E 的子集. 集合 F 的邻域是 E 的子集, 它是 F 中每一点的邻域. 点 $e \in F$ 称为 F 的内部点 (interior point), 若存在 e 的邻域 V, 满足 $V \subset F$. F 的内部点的全体是 (可能是空集的) 开集, 记为 $\text{int}(F)$. 称 $e \in E$ 是 $F \subset E$ 的极限点(limit point), 若 e 的每一邻域均至少含有一不同于 e 的点 $f \in F$. 集合 $F \subset E$ 的闭包 (closure) 是包

含 F 的所有闭集的交, 记为 $\mathrm{cl}(F)$. F 的闭包是闭集, 它是 F 与它的极限点的并集. 开集的一基本系统 (fundamental system) $\mathcal{G}'$ 是 $\mathcal{G}$ 的子集, 满足：若 $e \in E$, V 是 e 的邻域, 则 V 包含 $\mathcal{G}'$ 中的开集, 此开集包含 e. 若对某一固定的 $e \in E$, 上述性质成立, 则称 $\mathcal{G}'$ 是 e 处的基本系统 (或基). 注意到, 对给定的 E 的子集类 $\mathcal{O}$, 若 $\mathcal{O}$ 是基本系统, 则 $\cup\{O \in \mathcal{O}\} = E$, 且 $\mathcal{O}$ 中有限个元素的交仍是 $\mathcal{O}$ 中的元素. 相反地, 若 $\mathcal{O}$ 满足这两个条件, 则对于由 $\mathcal{O}$ 的元素的 (可能是无限的) 并构成的拓扑, $\mathcal{O}$ 是基本系统. 集合 $F \subset E$ 称为在 E 中是稠密的(dense), 若 $\mathrm{cl}(F) = E$. 空间 E 称为是可分的 (separable), 若它含有可数的稠密子集. Hausdorff 空间是拓扑满足下述的所谓的分离公理(separation axiom) 的拓扑空间$(E, \mathcal{G})$：

$$\forall (e_1, e_2) \in E \times E,\ e_1 \neq e_2,\ \exists (G_1, G_2) \in \mathcal{G} \times \mathcal{G} : e_1 \in G_1,\ e_2 \in G_2 \text{ 且 } G_1 \cap G_2 = \varnothing.$$

考虑序列 $\{e_n\} \subset E$(除非说明, 指标 n 属于 $\mathbb{N} := \{1, 2, \cdots\}$, 自然数集). 称 e 是 $\{e_n\}$ 的极限点(limit point), 若对 e 的任何邻域 N 及 $n \in \mathbb{N}$, 存在 $m > n$ 满足 $e_m \in N$. 称 $\{e_n\}$ 收敛到 (converge) $e \in E$ 或 e 是 $\{e_n\}$ 的极限, 记为 $e_n \to e$, 若对 e 的任何邻域 N, 存在 n_0(依赖于 N) 满足只要 $n > n_0$ 就有 $e_n \in N$. 显然, 序列的极限也是该序列的极限点. 若空间 E 是 Hausdorff 的, 则尽管 $\{e_n\}$ 可以有几个极限点, 它的极限却是唯一的. 若对给定的 $F \subset E$, $\{e_n\} \subset F$ 且 $e_n \to e$, 则 $e \in \mathrm{cl}(F)$. 应注意到, 这一结论的逆命题通常是不真的, 即并不是 $\mathrm{cl}(F)$ 的每一点都可以表示为 F 的点列的极限. 即并不是每一拓扑都可以以收敛序列的方式来描述. 然而, 对一类重要的拓扑空间, 这一性质是成立的. 即若 E 是 Hausdorff 空间, 每一点 $e \in E$ 均有可数基(countable base), 则每一 $\mathrm{cl}(F)$ 均可表示为 F 中的所有收敛序列之极限的并.

设 F 是拓扑空间 $(E, \mathcal{G})$ 的子集. F 的引导拓扑 (induced topology) 是 $\{G \cap F : G \in \mathcal{G}\}$, 即 F 中的开集为 E 中的开集与 F 的交集. 它也满足拓扑公理.

设 $f : X \to Y$ 是一映射, 其中 X 与 Y 是拓扑空间. 称 f 在点 $x \in X$ 处是连续的 (continuous), 若对 $f(x)$ 的每一邻域 V, 存在 x 的邻域 U 满足 $f(U) \subset V$; 若 f 在每一点 $x \in M$ 处是连续的, 则称它在 $M \subset X$ 上是连续的. 当 $M = X$ 时, 称 f 是连续的. f 的连续性的充分必要条件是 Y 中的任意开 (闭) 集的逆象是 X 中的开 (闭) 集.

2. *度量空间*

集合 E 上的度量是满足下述公理的函数 $\rho : E \times E \to \mathbb{R}$ (其中 $\mathbb{R}$ 记实数集), 即对 $e_1, e_2, e_3 \in E$, 有

(i) $\rho(e_1, e_2) \geqslant 0$, 且 $\rho(e_1, e_2) = 0$ 当且仅当 $e_1 = e_2$.

(ii) 对称性：$\rho(e_1, e_2) = \rho(e_2, e_1)$.

(iii) 三角不等式：$\rho(e_1,e_3)\leqslant\rho(e_1,e_2)+\rho(e_2,e_3)$.

令 r 是正数. 记 $B(e,r):=\{y\in E:\rho(e,y)<r\}$ 是开球(open ball), 或简言之, 中心是 e, 半径是 r 的球. 相关联的度量拓扑 (metric topology) 由这样的子集 $U\subset E$ 作开集：若 $e\in U$, 则 U 包含以 e 为中心的开球. 因此, 开球构成度量拓扑的开集的基本系统. 易见, 这一拓扑满足分离公理, 因此, 度量空间(metric space) 是 Hausdorff 空间.

对度量空间E, $e\in E$, 开球类 $B(e,r)$, r 是正的有理数, 形成点 e 处的可数基. 因此, 度量空间的拓扑可用收敛序列来刻画. 序列 $\{e_n\}\subset E$ 收敛到 e 当且仅当 $\rho(e_n,e)\to 0$, e 是集 $F\subset E$ 的极限点当且仅当存在序列 $\{e_n\}\subset F\setminus\{e\}$ 收敛到 e. 由三角不等式又得, 开球 $B(e,r)$ 的闭包包含在闭球(closed ball)

$$\overline{B}(e,r)=\{y\in E:\rho(e,y)\leqslant r\}$$

中.

度量空间E 的序列 $\{e_n\}$ 被称为 Cauchy 序列, 若 $\lim\limits_{m,n\to\infty}\rho(e_n,e_m)=0$, 即对每一 $\varepsilon>0$, 存在 $n_\varepsilon\in\mathbb{N}$ 满足若 $m,n>n_\varepsilon$, 则 $\rho(e_n,e_m)<\varepsilon$. Cauchy 序列至多有一个极限点, 任何收敛序列是 Cauchy 序列. 度量空间E 被称为是完备的 (complete), 若 E 中的每一 Cauchy 序列均是收敛的.

引理 2.1 (Baire 引理) *设 (E,ρ) 是完备的度量空间, $\{F_n\}$ 是 E 的闭子集序列. 若对所有的 $n\in\mathbb{N}$ 有 $\mathrm{int}(F_n)$ 均是空集, 则 $\mathrm{int}\Big(\bigcup\limits_{n\in\mathbb{N}}F_n\Big)$ 也是空集.*

证明 定义 $G_n:=E\setminus F_n$. 只需证明 $G:=\bigcap\limits_{n\in\mathbb{N}}G_n$ 在 E 中是稠密的, 即对任何 $x_0\in E$, $r_0>0$, 集合 $G\cap B(x_0,r_0)$ 非空. 由于 G_1 是开的, 且在 E 中是稠密的, 存在 $x_1\in G_1, r_1\in(0,r_0/2)$ 满足 $B(x_1,r_1)\subset G_1\cap B(x_0,r_0/2)$. 由归纳法得到序列 $x_n\in E$, $r_n>0$ 满足 $r_n<r_{n-1}/2$, 且

$$\overline{B}(x_n,r_n)\subset G_n\cap B(x_{n-1},r_{n-1}/2).$$

显然, $\{x_n\}$ 是 Cauchy 序列. 由于 E 是完备的. $\{x_n\}$ 有极限 $x\in E$, 它属于 $B(x_0,r_0/2)$ 的闭包, 因此属于 $B(x_0,r_0)$. 由构造, x 属于所有的 $G_n, n\in\mathbb{N}$. 证毕. □

映射 $T:E\to E$ 称为是压缩的 (contracting), 若存在常数 $\alpha\in(0,1)$, 满足

$$\rho(T(e_1),T(e_2))\leqslant\alpha\rho(e_1,e_2),\quad\forall\, e_1,e_2\in E.\tag{2.1}$$

定理 2.2 (不动点定理) *设 (E,ρ) 是完备的度量空间, $T:E\to E$ 是压缩映射. 则存在唯一点 $\bar{e}\in E$ 满足 $T(\bar{e})=\bar{e}$.*

证明 考虑 E 中如下定义的序列 $\{e_i\}$, 初始点为 $e_0\in E$, $e_{i+1}=T(e_i)$, $i=1,2,\cdots$. 这一序列是 Cauchy 列. 实际上, 对任何 $k\in\mathbb{N}$,

$$\rho(e_k,e_{k+1})=\rho(T^k(e_0),T^k(e_1))\leqslant\alpha^k\rho(e_0,e_1),$$

因此, 对任何 $n > m$,

$$\begin{aligned}\rho(e_m, e_n) &\leqslant \rho(e_m, e_{m+1}) + \cdots + \rho(e_{n-1}, e_n) \\ &\leqslant \alpha^m(1 + \alpha + \cdots + \alpha^{n-m-1}) \leqslant (1-\alpha)^{-1}\alpha^m.\end{aligned}$$

由于 $\{e_i\}$ 是 Cauchy 序列, 它收敛到一点 $\bar{e} \in E$. 由于压缩映射是连续的, 得到

$$\bar{e} = \lim_{i\to\infty} e_i = \lim_{i\to\infty} e_{i+1} = \lim_{i\to\infty} T(e_i) = T(\bar{e}),$$

即 $\bar{e}$ 是 T 的不动点. 不动点的唯一性由 (2.1) 立即得到. □

3. 紧致集合

设 F 是拓扑空间 E 的子集. F 的覆盖 (covering) 是 E 的一族子集, 它们的并包含 F, 即存在一族 $\{F_i\}_{i\in I} \subset 2^E$[①]满足 $F \subset \bigcup_{i\in I} F_i$. 覆盖 $\{F_i\}_{i\in I}$ 被称为是开的 (open), 若每一 F_i 均是开集, 被称为是有限的, 若 I 是有限集. $\{F_i\}_{i\in I}$ 中的 F 的子覆盖具有形式 $\{F_i\}_{i\in J}$, 其中 $J \subset I$. 称集合 $F \subset E$ 是紧致的 (compact), 若 F 的任意开覆盖均含有有限子覆盖. 注意到 F 是紧致的当且仅当它由 F 在 E 中引导的拓扑亦是紧致的.

命题 2.3 (i) 紧致集在连续映射下的像亦是紧致的.

(ii) Hausdorff 拓扑空间中的任何紧致子集是闭集.

(iii) 紧致空间的任何闭子集是紧致的.

证明 (i) 设 E 与 G 是两个拓扑空间, F 是 E 的紧致子集, $f: E \to G$ 是连续函数. 设 $\{G_i\}_{i\in I} \subset 2^G$[②]是 $f(F)$ 的开覆盖. 则由于 f 是连续的, $\{f^{-1}(G_i)\}_{i\in I}$ 是 F 的覆盖, 而且是开覆盖. 因为 F 是紧致的, 存在子覆盖 $\{f^{-1}(G_i)\}_{i\in J}$, J 是有限集. 可得到 $\{G_i\}_{i\in J}$ 是 $f(F)$ 的有限覆盖. 因此, $f(F)$ 是紧致的.

(ii) 设 F 是 Hausdorff 拓扑空间 $(E, \mathcal{G})$ 的紧致子集. 若 $e_0 \in E \setminus F$, 由分离公理, 对每一 $e \in F$, 存在 e 与 e_0 的开邻域 V_e 与 U_e, 它们的交集是空集. 由开覆盖 $\{V_e, e \in F\}$ 可抽取 F 的有限开覆盖 $\{V_e, e \in I\}$. 集合 $\bigcap_{e\in I} U_e$ 是包含在 $E \setminus F$ 中的 e_0 的一邻域. 因此, $E \setminus F$ 是开集, 即 F 是闭集.

(iii) 设 F 是紧致拓扑空间 E 的闭子集. 设 $\{V_i, i \in I\}$ 是 F 的开覆盖. 将 $E\setminus F$ 添加到此覆盖, 得到 E 的开覆盖. 由于 E 是紧致的, 可选取有限覆盖 $\{E \setminus F, V_i, i \in J\}$, J 是 I 的有限子集. 则 $\{V_i, i \in J\}$ 是 F 的有限覆盖. 因此, F 是紧致的. □

命题 2.4 拓扑空间 E 的子集 F 是紧致的充分必要条件是 F 的任何具有空集交的一族闭子集均含有具有空集交的有限子族.

---

① 原文为 $\{F_i\}_{i\in I} \subset E$.

② 原文为 $\{G_i\}_{i\in I} \subset G$.

证明 设 F 是 E 的紧致子集, $\{F_i\}_{i\in I}$ 是具有空集交的 F 的一族闭集. 则 $\{E\setminus F_i\}_{i\in I}$ 是 F 的开覆盖. 设 $\{E\setminus F_i\}_{i\in J}$ 是 F 的有限开覆盖. 则 $\{F_i\}_{i\in J}$ 是具有空集交的有限子族. 相反地, 设 F 是 E 中具有所述性质的子集. 令 $\{V_i, i\in I\}$ 是 F 的开覆盖. 则 $\{F\setminus V_i, i\in I\}$ 是具有空集交的一族闭子集. 令 $\{F\setminus V_i, i\in J\}$ 是具有空集交的有限子族. 则 $\{V_i, i\in J\}$ 是 F 的有限覆盖. 因此, F 是紧致的. □

上述命题表明, 紧致空间的任何序列至少有一极限点. 事实上, 设 $\{e_k\}$ 是紧致空间 E 的点列. 令 F_n 是集合 $\bigcup\limits_{k\geqslant n}\{e_k\}$ 的闭包. 则 F_n, $n=1,2,\cdots$ 是一非增的闭集合, 具有性质: 任何有限子类均有非空的交. 所以, $\bigcap\limits_{n=1}^{\infty}F_n$ 是非空的. 这就是 $\{e_k\}$ 的极限点的集合. 这一结果的一个结论是, 若 E 是 Hausdorff 空间且每一点 $e\in E$ 均有可数基, 则紧致空间的任何序列均含有收敛子序列.

拓扑空间 E 的子集 F 是序列紧的 (sequentially compact), 若每一序列 $\{e_n\}\subset F$ 均有收敛到 F 中的某一元素的子序列 $\{e_{n(k)}\}$. 由上述讨论可知, 若 E 是 Hausdorff 空间, E 中每一点均有可数基, 则 E 的每一紧致子集是序列紧的. 若 E 是度量空间, 则相反的结论也是成立的. 即度量空间是紧致的充分必要条件是它是序列紧的.

4. 增广实值函数

设 E 是拓扑空间, f 是定义在 E 上的增广实值函数, 即记 $f: E\to\overline{\mathbb{R}}$, $\overline{\mathbb{R}}:=\mathbb{R}\cup\{-\infty\}\cup\{+\infty\}$ 为增广实数轴. 集合

$$\mathrm{gph}f := \{(e,f(e)) : e\in E\},\quad \mathrm{epi}f := \{(e,\alpha)\in E\times\mathbb{R} : \alpha\geqslant f(e)\}$$

称为 f 的图 (graph) 与上图(epigraph). f 在 e 处的下极限 (lower limit) 是

$$\liminf_{e'\to e} f(e') := \sup_{V\in\mathcal{V}(e)} \inf\{f(e') : e'\in V, e'\neq e\},$$

其中 $\mathcal{V}(e)$ 是 e 的一个基. 序列 $f(e_n)$ 的下极限是

$$\liminf_{n\to\infty} f(e_n) := \lim_{n\to\infty}\inf_{k\geqslant n}\{f(e_k)\}.$$

上极限 (upper limit) 通过下述关系定义

$$\limsup f = -\liminf(-f).$$

称 f 在点 e 处是下半连续的 (lower semicontinuous), 若对任意的 $\varepsilon>0$, 存在 e 的邻域 U 满足 $f(U)\subset[f(e)-\varepsilon,+\infty]$. 等价地, f 在 e 处是下半连续的充分必要条件是

$$\liminf_{e'\to e} f(e')\geqslant f(e).$$

称 f 是下半连续的 (l.s.c.), 若它在每一点 $e \in E$ 处均是下半连续的. 称 f 是上半连续的 (u.s.c.), 若 $(-f)$ 是 l.s.c.. 若 E 是度量空间, 则其拓扑可由收敛序列来描述, 此种情况, f 在 e 处下半连续当且仅当对任何序列 $e_n \to e$, $f(e) \leqslant \liminf_{n\to\infty} f(e_n)$.

所有不超过 f 的下半连续函数 g (即对所有 $e \in E$, $g(e) \leqslant f(e)$) 的上确界函数被称为 f 的 l.s.c. 包 (l.s.c. hull), 记为 $\mathrm{lsc}f$. 定义 f 的水平集(level set)

$$\mathrm{lev}_{\leqslant\alpha} f := \{e \in E : f(e) \leqslant \alpha\},$$

其中 $\alpha \in \mathbb{R}$. 下述命题的结论容易由定义证得, 留作练习.

命题 2.5 (i) *若函数 f 是连续的, 则其图是闭的.*

(ii) *函数 f 是 l.s.c. 当且仅当其上图是闭的.*

(iii) *函数 f 是 l.s.c. 当且仅当它的所有水平集均是闭集.*

(iv) *令 $\{f_i\}_{i\in I}$ 是一族下半连续函数. 则最大值函数 $f(e) := \sup_{i\in I} f_i(e)$ 亦是下半连续函数.*

(v) *f 的上图的拓扑包是 $\mathrm{lsc}f$ 的上图.*

由 (v) 得, 函数 f 是下半连续的当且仅当 $\mathrm{lsc}f = f$. 由下半连续函数的定义, 有

$$\mathrm{lsc}f(e) = \min\{f(e), \liminf_{e'\to e} f(e')\}. \tag{2.2}$$

函数 f 的定义域 (domain) 定义为

$$\mathrm{dom}f := \{e \in E : f(e) < +\infty\}.$$

称函数 f(在 E 上) 获得最小值, 若集合

$$\operatorname*{argmin}_E f = \operatorname*{argmin}_{x\in E} f(x) := \{e \in \mathrm{dom}f : f(e) \leqslant f(e'),\ \forall e' \in E\}$$

是非空的. 注意到, 设 $e \in \mathrm{dom}f$, f 于 $\operatorname*{argmin}_E f$ 上可取 $-\infty$, 但不取 $+\infty$. f 在 E 上的最大值点集合定义为

$$\operatorname*{argmax}_E f := \operatorname*{argmin}_E (-f).$$

称 $f : E \to \overline{\mathbb{R}}$ 是下确界–紧致的 (inf-compact), 若对某一 $\alpha \in \mathbb{R}$, 水平集$\mathrm{lev}_{\leqslant\alpha} f$ 是非空的且紧致的.

定理 2.6 *设 $f : E \to \overline{\mathbb{R}}$ 是下确界–紧致的和下半连续的. 则 f 在 E 上取得最小值.*

证明 置 $v(f) := \inf_{e\in E} f(e)$. 因为存在某一 $\alpha \in \mathbb{R}$, 相应的水平集$\mathrm{lev}_{\leqslant\alpha} f$ 是非空的, 有 $v(f) < +\infty$. 若 $v(f)$ 是有限的且 $\alpha = v(f)$, 则 $\mathrm{lev}_{\leqslant\alpha} f$ 的任何元素均是 f 的极小值点, 结论得证. 因此, 设 $\alpha > v(f)$. 考虑下述水平集

$$F_k := \begin{cases} e \in E : f(e) \leqslant v(f) + 1/k, & \text{若 } v(f) \text{ 有限}, \\ e \in E : f(e) \leqslant -k, & \text{若 } v(f) = -\infty. \end{cases}$$

由 $v(f)$ 的定义有这些水平集是非空的, 由于 f 是 l.s.c., 它们是闭集, 对 k 充分大有它们是 $\text{lev}_{\leqslant\alpha} f$ 的子集. 进一步, 这族闭集合满足有限子族交非空的性质. 由命题 2.4 得, $\{F_k\}_{k\in\mathbb{N}}$ 的交是非空集. 这一交集即 f 取最小值的点集, 它是非空的. □

极小点的存在性经常采用下述的构造方法来证明. 在 E 上极小化增广实值函数 f 问题的一个极小化序列 (minimizing sequence) 定义为一个序列 $\{e_n\} \subset E$, 满足对所有的 n, $f(e_n) < +\infty$ 且 $f(e_n) \to \inf\{f(e) : e \in E\}$. 若定理 2.6 的假设成立, 则极小化序列是存在的, 且由于 f 的水平集非空紧致, 这一序列至少有一个极限点(见命题 2.4 后面的讨论), 尽管通常来讲不能从其抽出一收敛子列. 由 f 的下半连续性, 这一极限点是 f 在 E 上的最小点. 因此, 在紧致水平集的情况下, 基于极小化序列的证明本质上与定理 2.6 的证明相同.

有两类问题, 其极小点的存在性不需要紧致性假设就可以证明：凹与线性规划问题 (见定理 2.198) 及二次规划问题 (见定理 3.128). 2.5 节中给出的对偶理论也包含各种各样的存在性定理.

5. *拓扑向量空间*

集合 X 上的叠加律 "+" 是一映射 $X \times X \to X$, 记为 $(x_1, x_2) \mapsto x_1 + x_2$. 称 $(X, +)$ 是一群(group), 若下述三条公理成立：

(i) $+$ 是结合的 (associative), 即 $(x_1 + x_2) + x_3 = x_1 + (x_2 + x_3)$, $\forall x_1, x_2, x_3 \in X$.

(ii) 存在自然元素 (natural element) 记为 0, 即 $x + 0 = 0 + x = x$, $\forall x \in X$.

(iii) 每一元素 x 均有对称的元素 (symmetric element), 记为 $-x$, 满足 $x + (-x) = (-x) + x = 0$.

若对 $\forall x_1, x_2 \in X$, 还成立着 $x_1 + x_2 = x_2 + x_1$, 则群被称为是 Abel 的或交换的 (Abelian, commutative).

$\mathbb{R}$ 上的向量 (线性) 空间 (vector (linear) space) X 是 Abel(交换) 群, 其叠加律被称为加法 (addition), 且赋有从 $\mathbb{R} \times X$ 到 X 的映射 $(\alpha, x) \mapsto \alpha x$ 满足对所有的 $\alpha, \beta \in \mathbb{R}, x, y \in X$, 还有下述四条公理成立：

(i) $\alpha(x + y) = \alpha x + \alpha y$, (ii) $(\alpha + \beta)x = \alpha x + \beta x$,

(iii) $(\alpha\beta)x = \alpha(\beta x)$, (iv) $1x = x$.

注意, 本书不考虑 $\mathbb{R}$ 以外的域上的向量空间.

令 x_1 与 x_2 属于向量空间 X. 线段(segment)$[x_1, x_2]$ 定义为

$$[x_1, x_2] := \{\alpha x_1 + (1 - \alpha)x_2 : 0 \leqslant \alpha \leqslant 1\}. \tag{2.3}$$

向量空间 X 赋予拓扑, 若满足代数运算 $(x,y)\mapsto x+y$ 与 $(\alpha,x)\mapsto\alpha x$ 在对应的直积拓扑上是连续的, 则称之为拓扑向量空间(topological vector space). 用 X^* 记 X 的拓扑对偶(topological dual), 即所有连续的线性泛函或线性形式 $x^*:X\to\mathbb{R}$ 的向量空间. 通常把值 $x^*(x)$ 写成 $\langle x^*,x\rangle$. 由 $+$ 的连续性, X 中的子集 V 是 x 的邻域当且仅当 $V-x$ 是 0 的邻域.

注意到线性泛函 $x^*:X\to\mathbb{R}$ 在 X 上连续的充分必要条件是它在 0 点连续. 因此, 线性形式 x^* 是连续的当且仅当存在 X 中 0 的邻域 N 满足: 只要 $x\in N$, 就有 $|x^*(x)|\leqslant 1$.

设 X 是向量空间. 函数 $p:X\to\mathbb{R}_+$ 称为半范数(seminorm), 若它满足对所有的 $x,y\in X$ 及 $\alpha\in\mathbb{R}$, 下述公理成立:

(i) 次可加性: $p(x+y)\leqslant p(x)+p(y)$.

(ii) 齐次性: $p(\alpha x)=|\alpha|p(x)$. $\qquad$ (2.4)

于是集合

$$M:=\{x\in X:p(x)\leqslant 1\} \tag{2.5}$$

满足下述性质:

(i) 集合 M 是凸的 (convex), 即若 $x,y\in M,\alpha\in[0,1]$, 则 $\alpha x+(1-\alpha)y\in M$.

(ii) 集合 M 是均衡的 (balanced), 即若 $x\in M$ 且 $|\alpha|\leqslant 1$, 有 $\alpha x\in M$.

(iii) 集合 M 是吸收的 (absorbing), 即对任何 $x\in X$, 存在 $\alpha>0$ 满足 $\alpha x\in M$.

称集合 $M\subset X$ 是桶 (barrel) 集, 若它满足上述性质 (i)~(iii). 由 (ii) 得, 任何桶集均包含 $0\in X$.

定义 2.7　*称拓扑向量空间 X 是局部凸的拓扑向量空间, 若 $0\in X$ 的任何邻域均包含开的桶集且 X 的拓扑是 Harsdorff 的, 即满足分离公理.*

因此, 在局部凸的拓扑向量空间中, 开的桶集的平移集构成开集的基本系统.

凸的吸收的 (不必是均衡的) 集合 M 联系着 Minkowski 度规(gauge)

$$p_M(x):=\inf\{\alpha:\alpha>0,\alpha^{-1}x\in M\}. \tag{2.6}$$

相反地, 有

$$M=\{x\in X:p_M(x)\leqslant 1\}.$$

若 M 是桶集, 则 $p_M(\cdot)$ 是半范数, 反之亦成立. 一般地, 可能发生 $x\neq 0,p_M(x)=0$, 此种情形, 桶集M 包含直线 $\{tx:t\in\mathbb{R}\}$.

引理 2.8　*设 X 是局部凸的拓扑向量空间, $M\subset X$ 是具有非空内部的凸的吸收的集合. 则*

$$\operatorname{int}(M)=\{x\in X:p_M(x)<1\}. \tag{2.7}$$

证明 存在 $x_0 \in \text{int}(M)$, 即有 0 的邻域 $\mathcal{O}$ 满足 $x_0 + \mathcal{O} \subset M$. 因为 M 是吸收的, 存在某一 $\beta > 0$ 满足 $(-\beta x_0) \in M$, 因此有

$$\frac{\beta}{1+\beta}(x_0 + \mathcal{O}) + \frac{1}{1+\beta}(-\beta x_0) \subset M.$$

从而得到 $\beta(1+\beta)^{-1}\mathcal{O} \subset M$.

若 $x \in \text{int}(M)$, 则存在 $\varepsilon > 0$ 满足 $(1+\varepsilon)x \in M$, 因而有 $p_M(x) \leqslant (1+\varepsilon)^{-1} < 1$. 相反地, 若 $p_M(x) < 1$, 令 $\alpha \in (0,1)$ 满足 $\alpha^{-1}x \in M$, 则

$$x + (1-\alpha)\frac{\beta}{1+\beta}\mathcal{O} \subset x + (1-\alpha)M \subset \alpha M + (1-\alpha)M = M. \tag{2.8}$$

因为 $\mathcal{O}$ 是 0 的一邻域, 得 $x \in \text{int}(M)$. □

从满足下述分离公理的一族半范数 $\{p_i(\cdot)\}_{i\in I}$ 出发, 可以构造局部凸的拓扑向量空间的拓扑:

$$\text{对任何 } x \in X, x \neq 0, \text{存在 } i \in I \text{ 满足 } p_i(x) > 0. \tag{2.9}$$

考虑所有具有下述形式的集合构成的族 $\mathcal{G}'$,

$$G = \bigcap_{i\in J}\{x \in X : p_i(x) < \varepsilon_i\}, \tag{2.10}$$

其中 $J \subset I$ 是有限指标集, $\varepsilon_i > 0, \forall i \in J$. 将 $\mathcal{G}'$ 定义为 0 处的基本系统 (基), 开集的基本系统 $\mathcal{G}$ 可通过平移 $x + G, G \in \mathcal{G}', x \in X$ 得到. 于是得到 X 上的 Hausdorff 拓扑, 任何半范数 p_i, $i \in I$ 在这一拓扑下均是连续的, 因此 X 成为拓扑向量空间. 进一步, 每一集合 $G \in \mathcal{G}'$ 均是开的桶集, 从而 $(X, \mathcal{G})$ 是局部凸的拓扑向量空间.

若局部凸的拓扑向量空间 X 是有限维的, 赋予内积运算 $\langle\cdot,\cdot\rangle$, 其拓扑可由半范数 $p_i(\cdot) := |\langle x_i, \cdot\rangle|$ 定义, 其中 $x_1, \cdots, x_n$ 是 X 的一组基. 即任何 n 维的局部凸的拓扑向量空间与 $\mathbb{R}^n$ 是同构的 (isomorphic).

例 2.9 设 Ω 是非空集合, 考虑线性空间 $\mathbb{R}^\infty(\Omega)$, 它由实值函数 $x : \Omega \to \mathbb{R}$ 组成, 赋予自然的代数运算, 半范数族为 $p_\omega(x) := |x(\omega)|, \omega \in \Omega$. $\mathbb{R}^\infty(\Omega)$ 中的序列 $\{x_n\}$ 收敛到元素 x 当且仅当对每一 $\omega \in \Omega$ 均有 $x_n(\omega) \to x(\omega)$, 即 $\mathbb{R}^\infty(\Omega)$ 拓扑对应着逐点收敛.

令 $\Omega := \mathbb{N}$, X 是由有界序列 $x = (x_i)$ 构成的 $\mathbb{R}^\infty(\mathbb{N})$ 的线性子空间. 则集合

$$M := \{(x_i) \in X : |x_i| \leqslant 1, i = 1, 2, \cdots\}$$

是具有空集内部的桶集. 因此, 引理 2.8 的假设中 M 具有非空内部是本质性的.

2.1.2　Hahn-Banach 定理

设 X 是向量空间. 称函数 $p: X \to \mathbb{R}$ 是正齐次的 (positively homogeneous), 若

$$p(\alpha x) = \alpha p(x), \quad \forall \alpha \geqslant 0, x \in X.$$

下述定理讨论被一个次可加 (见 (2.4)) 与正齐次函数控制的线性泛函的延拓. 其证明可在几乎任何一本泛函分析教科书中均可找到.

定理 2.10 (Hahn-Banach)　*设 X 是向量空间, $p: X \to \mathbb{R}$ 是一次可加且正齐次的函数, L 是 X 的线性子空间, $\ell(\cdot)$ 是定义在 L 上的线性泛函. 设 p 在下述意义下控制 ℓ, 即对所有 $x \in L$, 有 $\ell(x) \leqslant p(x)$. 则存在定义于 X 的线性泛函 $\hat{\ell}$, 它是 ℓ 的延拓且被 p 控制. 即 $\hat{\ell}$ 满足*

$$\hat{\ell}(x) = \ell(x), \quad \forall x \in L, \quad 且 \quad \hat{\ell}(x) \leqslant p(x), \quad \forall x \in X.$$

后面将用到 Hahn-Banach 定理的下述结论.

命题 2.11　*设集合 X 是局部凸的拓扑向量空间, L 是 X 中具有相应的引导拓扑的线性子空间, 令 ℓ 是定义于 L 上的连续的线性泛函. 则 ℓ 可延拓为 X 上的连续的线性泛函.*

证明　因为 ℓ 是 L 上的连续泛函, 存在开桶集$M \subset X$, 满足对所有 $x \in M \cap L$, 有 $|\ell(x)| \leqslant 1$. 设 $p(x) := p_M(x)$ 是 M 的 Minkowski 度规函数. 则在 L 上 $|\ell|$ 不超过 p, 因而 ℓ 可延拓为定义于 X 上的线性泛函 $\hat{\ell}$, 满足 $|\hat{\ell}|$ 被 p 控制 (或不超过 p). 尤其, 对 $x \in M$, $|\hat{\ell}(x)| \leqslant 1$. 因为 M 是 0 的邻域, 得到 $\hat{l}$ 在 X 上是连续的. □

函数 $\|\cdot\|: X \to \mathbb{R}_+$ 称为 X 上的范数 (norm), 若对任何 $x, y \in X$, $\alpha \in \mathbb{R}$, 有

(i) $\|x\| = 0$ 当且仅当 $x = 0$, (ii) $\|x + y\| \leqslant \|x\| + \|y\|$, (iii) $\|\alpha x\| = |\alpha|\|x\|$.

即范数满足半范数的公理, 另外还满足公理 (i). 赋予范数 $\|\cdot\|$ 的向量空间被称为赋范空间 (normed space). 定义在赋范空间 X 上的线性泛函 x^* 是连续的当且仅当它在单位球上是有界的, 即若

$$\|x^*\|_* := \sup_{\|x\| \leqslant 1} \langle x^*, x \rangle \tag{2.11}$$

是有限数. (2.11) 中的函数 $\|\cdot\|_*$ 定义了在拓扑对偶空间 X^* 上的范数, 被称为范数$\|\cdot\|$ 的对偶. 通常省略 “$*$”, 用 $\|x^*\|$ 表示对偶范数. 显然, (2.11) 可推出广义的 Cauchy-Schwartz 不等式 (generalized Cauchy-Schwartz inequality)

$$|\langle x^*, x \rangle| \leqslant \|x\| \, \|x^*\|. \tag{2.12}$$

命题 2.12　*设 x 是赋范空间 X 中的一点. 则存在 $x^* \in X^*$, 满足 $\|x^*\| = 1$ 且 $\langle x^*, x \rangle = \|x\|$.*

证明 考虑线性空间 $L:=\{tx:t\in\mathbb{R}\}$, 令 $p(\cdot):=\|\cdot\|$. 在 L 上定义下述线性泛函 $\ell(tx)=t\|x\|$, $t\in\mathbb{R}$. 由 Hahn-Banach 定理得, 这一线性泛函可以延拓为 X 上定义的线性泛函 x^*, 满足 $|\langle x^*,x\rangle|\leqslant\|x\|$, $\forall x\in X$. 由构造有 $|\langle x^*,x\rangle|=\|x\|$, 证毕. □

设 S 与 T 是向量空间 X 的两个子集. 称线性泛函 $x^*:X\to\mathbb{R}$ 分离 (separate) S 与 T, 若 $\langle x^*,x\rangle\leqslant\langle x^*,y\rangle$, $\forall x\in S$, $\forall y\in T$. 称 x^* 强分离 (strongly separate) S 与 T, 若存在 $\varepsilon>0$ 满足 $\langle x^*,x\rangle+\varepsilon\leqslant\langle x^*,y\rangle$, $\forall x\in S$, $y\in T$.

定理 2.13 (第一分离定理) *设 X 是局部凸的拓扑向量空间, 令 S 与 T 是 X 的两个凸子集, 设 S 具有非空内部. 则存在 $x^*\in X^*\setminus\{0\}$, 分离 S 与 T 的充分必要条件是 $\operatorname{int}(S)\cap T=\varnothing$.*

证明 用反证法证 "必要性". 假设 $\operatorname{int}(S)\cap T$ 非空, 如它包含 x_0, 则存在开的桶集 V 满足 $\{x_0\}+V\subset S$. 若 x^* 分离 S 与 T, 则对所有的 $x\in S$, $\langle x^*,x-x_0\rangle\leqslant 0$, 由于 V 是 $S-\{x_0\}$ 的均衡集合, 则对所有的 $x\in V$ 有 $\langle x^*,x\rangle=0$. 再由 V 是吸收的, 有 $\forall x\in X$, $\langle x^*,x\rangle=0$, 即 $x^*=0$.

再证 "充分性", 先证 $T=\{x_0\}$ 的情形. 如有必要, 可以通过平移, 不妨设 S 与 T 满足 $0\in\operatorname{int}(S)$. 设 x^* 是定义在 $\{tx_0:t\in\mathbb{R}\}$ 上的线性泛函, $x^*(tx_0)=t$, $t\in\mathbb{R}$. 令 $p(\cdot)$ 是与 S 相联系的 Minkowski 度规. 下证 $x^*(x)\leqslant p(x)$, $\forall x\in\{tx_0:t\in\mathbb{R}\}$. 事实上, 由引理 2.8 有, $p(x)<1$ 当且仅当 $x\in\operatorname{int}S$, 因此有 $x^*(x_0)=1\leqslant p(x_0)$. 由正齐次性可得

$$x^*(tx_0)=tx^*(x_0)\leqslant tp(x_0)=p(tx_0),\quad t\in\mathbb{R}_+.$$

最后, $x^*(tx_0)\leqslant 0\leqslant p(tx_0)$, 若 $t\in\mathbb{R}_-$, 可有 $x^*(x)\leqslant p(x)$, $\forall x\in\{tx_0:t\in\mathbb{R}\}$. 应用 Hahn-Banach 定理, 存在定义于 X 上的线性泛函 x^*, x^* 受控于 p (因而 x^* 是连续的). 因此

$$x^*(x)\leqslant p(x)\leqslant 1=x^*(x_0),\quad \forall x\in S,$$

从而 x^* 分离 S 与 $\{x_0\}$.

现在处理 T 是任意凸集的情形. 考虑集合 $C:=\operatorname{int}(S)-T=\bigcup_{y\in T}\{\operatorname{int}(S)-y\}$. 集合 C 是凸的, 由于 $\operatorname{int}(C)\supset\operatorname{int}(S)-T$, C 具有非空内部. 因 $\operatorname{int}(S)\cap T=\varnothing$, C 不含有 0. 因此, 集合 C 与 $\{0\}$ 可以被分离, 即存在非零的线性泛函 x^* 满足对所有的 $x\in C$ 有 $x^*(x)\leqslant 0$. 显然, x^* 分离 S 与 T. □

定理 2.14 (第二分离定理) *设 X 是局部凸的拓扑向量空间, S 与 T 是 X 的闭凸子集且 T 是紧致的, $S\cap T=\varnothing$. 则 S 与 T 可以被强分离.*

证明 考虑集合 $C:=S-T$. C 是凸的, 由 T 是紧致的, C 还是闭集. 事实上, 考虑一点 $x_0\notin C$. 由于 $C=\bigcup_{x\in T}(S-x)$, 这表明对任何 $x\in T$ 有 $x_0\notin S-x$. 由于 S 是闭的, $S-x$ 亦是闭的, 存在 $0\in X$ 的开桶集 V 满足 $(x_0+V)\cap(S-x_0)=\varnothing$,

或等价地, $x_0+\frac{1}{2}V\not\subset S-\left(x+\frac{1}{2}V\right)$ ①. 因为 T 是紧致的, 存在有限个点 $x_i\in T$ 与 $0\in X$ 的邻域 V_i, $i=1,\cdots,m$, 满足 $T\subset\bigcup_{i=1}^{m}\left(x_i+\frac{1}{2}V_i\right)$. 令 $V':=\frac{1}{2}\bigcap_{i=1}^{m}V_i$, 有 V' 是 $0\in X$ 的一邻域, 且

$$(x_0+V')\cap(S-T)\subset(x_0+V')\cap\left[\bigcup_{i=1}^{m}\left(x_i+\frac{1}{2}V_i\right)\right]=\varnothing.$$

这表明 C 是闭集.

由于 C 是闭集, 不包含 $0\in X$, 存在 0 的凸开集 V 满足 $V\cap C=\varnothing$. 由第一分离定理得, 存在非零的 $x^*\in X^*$ 分离 C 与 $\{0\}$, 即

$$\sup_{x\in C}x^*(x)\leqslant\inf_{y\in V}x^*(y).$$

于是存在, $\varepsilon>0$, 满足对所有的 $x\in C$, 均有 $x^*(x)\leqslant-\varepsilon$. 或等价地, 对任何 $x\in S$ 与 $y\in T$, $x^*(x)-x^*(y)\leqslant-\varepsilon$. 即 x^* 强分离 S 与 T, 证毕. □

注意, 尤其当 S 是空间 X 的闭凸子集且 $x_0\notin S$ 时, 上述定理表明, S 与 $\{x_0\}$ 可以被强分离.

推论 2.15 *局部凸的拓扑向量空间中的闭凸子集是包含它的闭半空间的交集.*

对拓扑向量空间 X 的子集 S, 用 $\mathrm{Sp}\,(S)$ 表示由 S 张成 (生成) 的线性空间, 即 $\mathrm{Sp}\,(S)$ 由形式为 $\alpha(x_1-x_2)$, $x_1,x_2\in S$, $\alpha\in\mathbb{R}$ 的元素构成的.

定义 2.16 *对拓扑向量空间 X 的凸子集 S, 称点属于 S 的相对内部, 记为 $\mathrm{ri}(S)$, 若 x 是 S 相对于由 S 生成的闭空间 $L:=\mathrm{cl}\{\mathrm{Sp}\,(S)\}$ 的内点. 即 $x\in\mathrm{ri}(S)$ 当且仅当存在 x 的在 X 中的邻域 N, 满足 $N\cap(x+L)\subset S$.*

若空间 X 是有限维的, 则 X 的每一线性子空间是闭的, 且 X 中的每一非空凸集均有非空的相对内部. 在无穷维空间的情形, 凸集的相对内部可能是空集. 如 $X:=L_2[0,1]$, $K\subset L_2[0,1]$ 是几乎处处非正值的函数的全体. 集合 K 是闭凸锥且 $\mathrm{Sp}\,(K)=L_2[0,1]$. 另一方面, K 在 $L_2[0,1]$ 中的内部是空集.

定理 2.17 *设 X 是局部凸的拓扑向量空间, S 是具有非空相对内部的 X 的凸子集. 令 x_0 是 X 中的一点, 满足 $x_0\notin\mathrm{ri}(S)$. 则 S 与 $\{x_0\}$ 可被分离.*

证明 令 $L:=\mathrm{cl}\{\mathrm{Sp}\,(S)\}$. 若 $x_0\notin L$, 则由定理 2.14, $\{x_0\}$ 与 L 可以被分离, 因此 S 与 $\{x_0\}$ 可被分离. 若 $x_0\in L$, $x_0\notin\mathrm{ri}(S)$, 则由定理 2.13, $S-x_0$ 与 x_0 可被定义于空间 L 上的线性的连续泛函分离. 这一线性泛函可被延拓为定义于空间 X 上的线性的连续泛函 (由命题 2.11), 结论得证. □

这里有 Hahn-Banach 定理的另一个有趣的结论.

① 原文为 $x_0+\frac{1}{2}V\notin S-\left(x+\frac{1}{2}V\right)$.

命题 2.18 *局部凸的拓扑向量空间 X 的有限维线性子空间 Z 是闭的.*

证明 若 $Z = X$, 则结论显然成立. 因此, 设 Z 是 X 的真子空间. 令 $x_1, \cdots, x_n$ 是 Z 的一组基, 设 $x_{n+1} \notin Z$. 令 $x_i^* \in X^*$ 满足 $\langle x_i^*, x_i \rangle \neq 0, 1 \leqslant i \leqslant n+1$. 置

$$M := \{x \in X: \ |\langle x_i^*, x_i \rangle| \leqslant 1, i = 1, \cdots, n+1\},$$

令 $p_M(\cdot)$ 是相应的 Minkowski 度规. 令 W 是由 $x_1, \cdots, x_{n+1}$ 生成的线性空间. 因为对 $y \in W$, $p_M(y) = 0$ 当且仅当 $y = 0$. p_M 限制到 W 是范数. 记 $y = \sum_{i=1}^{n+1} \alpha_i x_i$. 定义线性泛函 $\ell: W \to \mathbb{R}$, $\ell(y) = \alpha_{n+1}$. 由于 W 是有限维空间, p_M 是范数, 存在 $\beta > 0$, 对所有的 $x \in W$, $|\ell(x)| \leqslant \beta p_M(x)$. 由 Hahn-Banach 定理得, 这一泛函可被延拓为一连续的线性泛函 $\hat{\ell} \in X^*$. 因为 $\hat{\ell}$ 分离 x_{n+1} 与 Z, Z 可表示为一族闭的半空间的交, 因此它是闭的. □

设 S 是向量空间 X 的子集. S 的凸包(convex hull) 是 S 的所有有限元素凸组合的集合, 即

$$\operatorname{conv}(S) := \left\{ \sum_{i \in I} \alpha_i x_i : \ \alpha_i \geqslant 0, x_i \in S, \forall i \in I; \sum_{i \in I} \alpha_i = 1, |I| < \infty \right\}, \tag{2.13}$$

其中 “$|I| < \infty$” 意味着集合 I 是有限的. S 的凸包是包含 S 的最小凸集. 点 $x \in S$ 称为 S 的一个极点(extrme point), 若不存在不同于 x 的点 $x_1, x_2 \in S$, 使得 x 属于区间 $[x_1, x_2]$. 换言之, x 是 S 的极点, 若给定 $t \in (0,1)$, $x_1, x_2 \in S$ 满足 $x = tx_1 + (1-t)x_2$, 则 $x_1 = x_2 = x$.

定理 2.19 (Krein-Milman) *设 S 是局部凸的拓扑向量空间 X 的非空凸紧致子集, E 是 S 的所有极点的集合. 则 S 是 E 的凸包的拓扑包.*

证明 此定理基于下述结果, 它是 Zorn 引理的结论: 局部凸的拓扑向量空间的非空凸紧致子集至少有一个极点.

考虑集合 $C := \operatorname{cl}(\operatorname{conv} E)$. 它是紧致集 S 的闭子集, C 是紧致的 (命题 2.3). 若存在 $x \in S \setminus C$, 则由第二分离定理 2.14, 存在 $x^* \in X^* \setminus \{0\}$ 满足 $x^*(x) > \sup\{x^*(e) : e \in E\}$. 令 $S_1 := \operatorname{argmax}\{x^*(e) : e \in S\}$. 集合 S_1 是非空闭凸的, 它与 E 的交是空集. 由于 $S_1 \subset S$, S_1 是紧致的, 由上述结论, S_1 有一个极点. 由 S 的定义, 后者亦是 S 的极点, 它在 E 集之外, 这得到矛盾. □

注意, 若 S 是有限维空间 X 的非空凸紧致集合, 则 S 即是其极点集的凸包, 即在有限维情形, 上述定理中 E 的凸包不必再取拓扑包.

2.1.3 Banach 空间

完备的赋范空间 X 被称为 Banach 空间. 注意到, 任何赋范 (Banach) 空间 X, 赋予其范数拓扑, 是局部凸的拓扑向量空间. 因此, 由命题 2.11 有, 若 X_1 是赋范

(Banach) 实空间 X 的线性子空间, x_1^* 是定义于 X_1 上连续的线性泛函, 则 x_1^* 可以延拓为 X 上的连续的线性泛函, 即存在 $x^* \in X^*$, 满足 $\langle x^*, x_1 \rangle = \langle x_1^*, x_1 \rangle$, $\forall x_1 \in X_1$.

这一结果可如下表述. 令 i 记 X_1 到 X 的自然嵌入, 即 $i:\ X_1 \to X$, $i(x_1) = x_1$, $\forall x_1 \in X_1$. 伴随算子$i^*:\ X^* \to X_1^*$ 定义为

$$\langle i^* x^*, x_1 \rangle = \langle x^*, x_1 \rangle, \quad \forall x_1 \in X_1,$$

即 $i^* x^*$ 是线性形式 x^* 在 X_1 上的限制. 考虑下述图表, 其中映射 d 联系着 Banach 空间与其拓扑对偶. X_1^* 的任何元素可延拓到 X^* 意味着图表是可交换的 (commutative), 即 $i^* \circ d \circ i(X_1) = d(X_1) = X_1^*$.

$$\begin{array}{ccc} X & \xrightarrow{d} & X^* \\ {\scriptstyle i}\uparrow & & \downarrow{\scriptstyle i^*} \\ X_1 & \xrightarrow{d} & X_1^* \end{array}$$

若 Y 是另一 Banach 空间, 线性连续映射$A: X \to Y$ 的集合, 记为 $\mathcal{L}(X,Y)$, 它是赋予下述范数的 Banach 空间

$$\|A\| = \sup_{\|x\| \leqslant 1} \|Ax\|.$$

集合 $\mathcal{L}(X,X)$ 记为 $\mathcal{L}(X)$. 下述结果即所谓的一致有界原理 (principle of uniform boundedness).

定理 2.20 (Banach-Steinhaus) 设 X 与 Y 是 Banach 空间, $\mathcal{A} \subset \mathcal{L}(X,Y)$ 满足 $\sup_{T\in\mathcal{A}} \|Tx\| < \infty$, $\forall x \in X$. 则 $\sup_{T\in\mathcal{A}} \|T\| < \infty$.

证明 集合

$$X_n := \{x \in X : \|Tx\| \leqslant n, \forall T \in \mathcal{A}\}$$

是闭的, 由定理的假设有 $\bigcup_{n\in\mathbb{N}} X_n = X$. 则由 Baire 引理 2.1 得, 存在某一指标 n_0, 满足 $\operatorname{int}(X_{n_0}) \neq \varnothing$, 即 X_{n_0} 含有 $B(x_0, r_0)$, $x_0 \in X$, $r_0 > 0$. 因为 X_{n_0} 是凸的、均衡的, 有 $B(0, r_0) \subset X_{n_0}$, 即对任意 $T \in \mathcal{A}$, 只要 $\|x\| \leqslant r_0$, 就有 $\|Tx\| \leqslant n_0$, 或等价地, $\|T\| \leqslant n_0/r_0$. □

设 X 是 Banach 空间, 其拓扑对偶空间X^*(线性连续泛函 $x^*: X \to \mathbb{R}$ 的全体, 赋予 (2.11) 定义的对偶范数), 当赋予对偶范数时, 是 Banach 空间. 由命题 2.12, 因为 $\langle x^*, x \rangle \leqslant \|x^*\|\,\|x\|$, 对任何 $x \in X$, 有

$$\|x\| = \sup\{\langle x^*, x \rangle : \|x^*\| \leqslant 1, x^* \in X^*\}, \tag{2.14}$$

且其上确界是可达的. 称 $x \mapsto \underset{\|x^*\|\leqslant 1}{\operatorname{argmax}}\langle x^*, x\rangle$ 为对偶映射 (duality mapping)(通常这一映射是多值映射).

以类似的方式, 可定义拓扑双对偶 (topological bidual) X^{**} 为对偶的对偶, 即定义于对偶空间上的连续线性泛函 $x^{**}: X^* \to \mathbb{R}$ 构成的向量空间 (对偶空间赋予对偶范数的拓扑), 我们可给 X^{**} 赋予范数

$$\|x^{**}\| := \sup_{\|x^*\|\leqslant 1} \langle x^{**}, x^*\rangle.$$

任何 $x \in X$ 通过下述关系

$$\langle x^{**}, x^*\rangle := \langle x^*, x\rangle, \quad \forall x^* \in X^*$$

联系着 $x^{**} \in X^{**}$. 由 (2.14), 相应的映射 $x \mapsto x^{**}$ 是等距的 (isometric)(即 $\|x^{**}\| = \|x\|$), 且使我们可以把 X 与 X^{**} 的一闭子空间相等同. 若 $X = X^{**}$, 则称 X 是自反的(reflexive).

下述结果对有限多个线性方程附加解的范数限制的可解性给出刻画.

引理 2.21 (Helly 引理)　*设 $x_1^*, \cdots, x_p^*$ 是定义在 Banach 空间 X 上线性的连续泛函, $b \in \mathbb{R}^p, \alpha > 0$. 对每一 $\varepsilon > 0$, 存在 $x_\varepsilon \in X$, 满足*

$$\langle x_i^*, x_\varepsilon\rangle = b_i, \quad i = 1, \cdots, p, \quad \|x_\varepsilon\| \leqslant \alpha + \varepsilon \tag{2.15}$$

的必要与充分条件是对任意的 $\lambda \in \mathbb{R}^p$, 下述不等式成立

$$\left|\sum_{i=1}^p \lambda_i b_i\right| \leqslant \alpha \left\|\sum_{i=1}^p \lambda_i x_i^*\right\|. \tag{2.16}$$

证明　由对偶范数的定义, 必要性是显然的, 下证明充分性. 令 $M: X \to \mathbb{R}^p$ 是线性映射, $M(\cdot) := (\langle x_1^*, \cdot\rangle, \cdots, \langle x_p^*, \cdot\rangle)$. 则 (2.15) 不成立当且仅当凸集 $M[B(0, \alpha+\varepsilon)]$ 不包含 b. 由于 $M[B(0, \alpha+\varepsilon)]$ 是 $\mathbb{R}^p$ 的子集, 其相对内部非空. 由定理 2.17 知, $M[B(0, \alpha+\varepsilon)]$ 可以与 b 分离, 即存在 $\lambda_i \in \mathbb{R}^p$, 满足

$$\sum_{i=1}^p \lambda_i b_i \geqslant \sum_{i=1}^p \lambda_i \langle x_i^*, x\rangle, \quad \forall x \in B(0, \alpha+\varepsilon).$$

在上面不等式之右端关于 $x \in B(0, \alpha+\varepsilon)$ 取最大值, 得到

$$\sum_{i=1}^p \lambda_i b_i \geqslant (\alpha+\varepsilon) \left\|\sum_{i=1}^p \lambda_i x_i^*\right\|,$$

这与 (2.16) 相矛盾. 结论得证. □

下述引理在研究无穷维线性规划时被用到, 见定理 2.204 证明.

引理 2.22 设 X 是 Banach 空间, $x_i^* \in X^*, i=1,\cdots,p, x^{**} \in X^{**}$. 则对任意 $\varepsilon > 0$, 存在 $x_\varepsilon \in X$ 满足 $\|x_\varepsilon\| \leqslant \|x^{**}\| + \varepsilon$ 且 $\langle x_i^*, x_\varepsilon \rangle = \langle x^{**}, x_i^* \rangle, i=1,\cdots,p$.

证明 令 $\alpha := \|x^{**}\|, b_i := \langle x^{**}, x_i^* \rangle, i=1,\cdots,p$. 对任何 $\lambda \in \mathbb{R}^p$,

$$\left|\sum_{i=1}^p \lambda_i b_i\right| = \left|\left\langle x^{**}, \sum_{i=1}^p \lambda_i x_i^* \right\rangle\right| \leqslant \alpha \left\|\sum_{i=1}^p \lambda_i x_i^*\right\|.$$

由 Helly 引理2.21 可得结论. □

在局部凸的拓扑向量空间X, 考虑由半范数族 $p_{x^*}(\cdot) := |\langle x^*, \cdot \rangle|, x^* \in X^*$ 定义的拓扑. 等价地, 这一拓扑由

$$G := \{x \in X : |\langle x_i^*, x \rangle| < \varepsilon_i, i=1,\cdots,n\}$$

构成的 0 的邻域的基本系统来定义, 其中 ε_i 是正常数. 由定理 2.14, X 中的任何两个不同点均可被强分离, 因此上述拓扑是 Hausdorff 拓扑. 此拓扑也称为 X 的弱拓扑(weak topology), 在此拓扑中, 每一泛函 $x^* \in X^*$ 是连续的. 若序列 $\{x_n\} \subset X$ 以弱拓扑收敛到一点 $\bar{x}$, 则称 x_n 弱收敛到 (weakly converges) $\bar{x}$, 记为 $x_n \xrightarrow{w} \bar{x}$. 不难看到, $x_n \xrightarrow{w} \bar{x}$ 当且仅当对任意的 $x^* \in X^*$, 均有 $\langle x^*, x_n \rangle \to \langle x^*, \bar{x} \rangle$. 注意, 除非空间 X 是有限维的, X 中的弱拓扑在 0 处没有可数的基.

我们也可以在对偶空间 X^* 中赋予所谓的弱星 (weak star)(弱 *) 拓扑, 它由半范数 $p_x(\cdot) := |\langle \cdot, x \rangle|, \forall x \in X$ 来定义 (不要同 X^* 的弱拓扑相混淆, 后者由半范数 $p_{x^{**}}(\cdot) := |\langle x^{**}, \cdot \rangle|, x^{**} \in X^{**}$ 定义). 由线性泛函的定义, 这一拓扑是 Hausdorff 的, 0 处的开集基本系统由

$$\{x^* \in X^* : |\langle x^*, x_i \rangle| \leqslant \varepsilon_i, i=1,\cdots,n\}$$

构成, 其中 ε_i 是正常数. 序列 $\{x_n^*\} \subset X^*$ 弱 * 收敛到 $\bar{x}^*$(记为 $x_n^* \xrightarrow{w^*} \bar{x}^*$) 当且仅当对任意 $x \in X$, 有 $\langle x_n^*, x \rangle \to \langle \bar{x}^*, x \rangle$.

定理 2.23 设 $\{x_n\}$ 是 Banach 空间X 的序列. 则

(i) 若 $x_n \to \bar{x}$, 则 $x_n \xrightarrow{w} \bar{x}$.

(ii) X 的一凸子集是弱闭的当且仅当它是强闭的.

(iii) 若 $x_n \xrightarrow{w} \bar{x}$, 则 $\|x_n\|$ 是有界的, 且 $\|\bar{x}\| \leqslant \liminf\limits_{n\to\infty} \|x_n\|$.

(iv) 若 $x_n \xrightarrow{w} \bar{x}, x_n^* \to \bar{x}^*$, 则 $\langle x_n^*, x_n \rangle \to \langle \bar{x}^*, \bar{x} \rangle$.

证明 (i) 若 $x_n \to \bar{x}$ 且 $x^* \in X^*$, 则 $|\langle x^*, x_n - \bar{x} \rangle| \leqslant \|x^*\| \, \|x_n - \bar{x}\| \to 0$, 因此 $\langle x^*, x_n \rangle \to \langle x^*, \bar{x} \rangle$. 这就证得 $x_n \xrightarrow{w} \bar{x}$.

(ii) 由第二分离定理 2.14, 强闭凸子集是包含它的半空间的交. 显然, 半空间是弱闭子集. 任何闭集合的交均是闭集, 强闭凸子集也是弱凸闭的.

另一方面, 某一点 x_0 的任何弱邻域包含 x_0 的强邻域. 因此弱开集是强开集. 因为闭集是开集的补集, 得到弱闭集是强闭的. 结论 (ii) 成立.

(iii) 若 $x_n \xrightarrow{w} \bar{x}$, 则对所有 $x^* \in X, \langle x^*, x_n\rangle$ 是有界的. 由 Banach-Steinhaus 定理 (定理 2.20) 得 $\|x_n\|$ 是有界的. 如有必要, 可选出一子列, 假设 $r := \lim\limits_{n\to\infty}\|x_n\|$ 存在. 固定 $\varepsilon > 0$. 对充分大的 n, x_n 属于闭凸集 $\overline{B}(0, r+\varepsilon)$. 由结论 (ii), 其极限属于 $\overline{B}(0, r+\varepsilon)$, 因此 (由于 $\varepsilon > 0$ 是有限的) 属于 $\overline{B}(0, r)$. 结论 (iii) 得证.

(iv) 若 $x_n \xrightarrow{w} \bar{x}$ 且 $x_n^* \to \bar{x}^*$, 由 (iii), $r := \sup\|x_n\| < \infty$. 所以有 $|\langle x_n^* - x^*, x_n\rangle| \leqslant r\|x_n^* - x^*\| \to 0$. 由于 $\langle \bar{x}^*, x_n\rangle \to \langle \bar{x}^*, \bar{x}\rangle$, 结果得证. □

相类似地, 可证明下述命题.

命题 2.24 *设 X 是 Banach 空间, $\{x_n^*\}$ 是 X^* 的序列. 则*

(i) 若 $x_n^* \to \bar{x}^*$, 则 $x_n^* \xrightarrow{w^*} \bar{x}^*$.

(ii) 若 $x_n^* \xrightarrow{w^*} \bar{x}^*$, 则 $\|x_n^*\|$ 是有界的且 $\|\bar{x}^*\| \leqslant \liminf \|x_n^*\|$.

(iii) 若 $x_n \to \bar{x}, x_n^* \xrightarrow{w^*} \bar{x}^*$, 则 $\langle x_n^*, x_n\rangle \to \langle \bar{x}^*, \bar{x}\rangle$.

下述定理表明, Banach 空间X 可视为 X^* 的拓扑对偶, 若后者赋予弱 * 拓扑.

定理 2.25 *令 X 是 Banach 空间, f 是 X^* 上的线性泛函, 它以弱 * 拓扑是连续的. 则存在 $x \in X$ 满足 $f(x^*) = \langle x^*, x\rangle, \forall x^* \in X^*$. 即 X^* 的对偶赋予弱 * 拓扑与 X 是同构的.*

证明 设 f 是 X^* 上的线性形式, 当 X^* 赋予弱 * 拓扑时是连续的. 则 f 在 0 的弱 * 邻域上是有界的, 即存在 $x_1, \cdots, x_n \in X, \alpha > 0$ 及 $\beta > 0$, 满足若 $|\langle x^*, x_i\rangle| \leqslant \beta, i = 1, \cdots, n$, 则 $f(x^*) \leqslant \alpha$. 考虑从 X^* 到 $\mathbb{R}^n$ 的线性映射 $L(x^*) := (\langle x^*, x_1\rangle, \cdots, \langle x^*, x_n\rangle)$. 由 $L(x^*) = 0$ 可得 $f(x^*) = 0$. 这意味着只要 $L(x_1^*) = L(x_2^*)$, 就有 $f(x_1^*) = f(x_2^*)$, 因此, $f(x^*)$ 是 $L(x^*)$ 的函数, 即 $f(x^*) = F(L(x^*)), \forall x^* \in X^*$. 映射 F 是从 $L(X^*) \subset \mathbb{R}^n$ 到 $\mathbb{R}$ 的线性映射. 将 F 在整个 $\mathbb{R}^n$ 上的线性拓延也记为 F. 将 F 在 $\mathbb{R}^n$ 中一组基分解得到 $F(y) = \sum\limits_{i=1}^{n} \gamma_i y_i, y \in \mathbb{R}^n$, 因此有 $f(x^*) =$

$$\sum_{i=1}^{n} \gamma_i \langle x^*, x_i\rangle = \left\langle x^*, \sum_{i=1}^{n} \gamma_i x_i \right\rangle.$$ □

上述定理启发我们引入下述概念.

定义 2.26 *设 X 与 Y 是局部凸的拓扑向量空间, 令 $\langle\cdot,\cdot\rangle$ 是定义于 $X \times Y$ 上的线性形式, 即对于每一 $y \in Y, \langle\cdot, y\rangle$ 是定义在 X 上的线性泛函, 对每一 $x \in X, \langle x, \cdot\rangle$ 是定义于 Y 上的线性泛函. 对 $\langle\cdot,\cdot\rangle$, 称 X 与 Y 有相容的拓扑, 或简言之, X 与 Y 是成对的空间(paired spaces), 若 X 上定义的线性连续泛函的全体与集合 $\{\langle\cdot, y\rangle : y \in Y\}$ 相同, Y 上定义的线性连续泛函的全体与集合 $\{\langle x, \cdot\rangle : x \in X\}$ 相同.*

由上述定义得到, 空间 Y 的任何元素 y 可等同于对偶空间 X^* 的元素 x^*, 类似

地, 通过关系 $x^*(\cdot) = \langle \cdot, y \rangle$ 联系, Y^* 的每一元素等同于 X 的某一元素.

由定理 2.25, Banach 空间 X 与其对偶是成对空间, 若 $X(X^*)$ 赋予强 (弱 *) 拓扑. 用定理 2.25 的结论有, 当 X 赋予其弱拓扑时, 其对偶还是 X^*. 因此, 另一个成对空间的例子是当 $X(X^*)$ 赋予弱 (弱 *) 拓扑时. 当然, 若 X 是自反的, 则 (X, X^*) 是成对空间, 如果它们被赋予强拓扑.

下述定理是 Tychonoff 定理 (即紧致集合的任意积是紧致集) 的一个结论.

定理 2.27 (Banach-Alaoglu) *设 X 是 Banach 空间. 则 X^* 的闭的单位球是弱 * 紧致的.*

当然, X^* 中的任何闭球是弱 * 紧致的. 结果, X^* 的任何有界的 (以 X^* 的范数) 弱 * 闭子集是弱 * 紧致的. 注意到, 由于自反的 Banach 空间 X 是其对偶空间的对偶 (二者均赋予强拓扑), 所以自反 Banach 空间的闭单位球是弱紧致的.

定理 2.28 (i) *Banach 空间是自反的当且仅当从任何有界序列中都可以选取弱收敛的子序列.*

(ii) *设 X 是可分的 Banach 空间. 则从 X^* 的任何有界序列中均可选取弱 * 收敛的子序列.*

结合定理 2.27 与 2.6 的结果, 可得下述推论.

推论 2.29 *设 X 是 Banach 空间. 则*

(i) *若 $f : X^* \to \overline{\mathbb{R}}$ 是弱 * l.s.c. 函数 (即按 X^* 的弱 * 拓扑是 l.s.c.), 满足至少有一水平集是非空有界的, 则 f 于 X^* 上取得最小值.*

(ii) *若空间 X 是自反的且 $f : X \to \overline{\mathbb{R}}$ 是弱 l.s.c. 函数, 满足至少有一水平集是非空有界的, 则 f 于 X 可得到最小值.*

注意到, 在上述推论中, 函数 f 是增广实值函数, 总可以添加约束 $x \in S$ 到极小化问题中, 通过加指示函数 $I_S(\cdot)$ 到函数 f 上即可.

例 2.30 (凸函数) 向量空间 X 上定义的增广实值函数 $f(x)$ 被称为是正常的 (proper), 若其 (有效) 定义域非空, 且对所有 $x \in X$, 有 $f(x) > -\infty$. 称 f 是凸的, 若对任何 $x_1, x_2 \in \operatorname{dom} f, t \in [0, 1]$,

$$f(tx_1 + (1-t)x_2) \leqslant tf(x_1) + (1-t)f(x_2). \tag{2.17}$$

等价地, f 是凸的, 若其上图

$$\operatorname{epi} f := \{(x, c) \in X \times \mathbb{R} : f(x) \leqslant c\} \tag{2.18}$$

是 $X \times \mathbb{R}$ 的凸子集. 由命题 2.5, X 上的函数是 l.s.c.(以 X 的强拓扑) 的充分必要条件是其上图是一强闭集, 类似的结论对弱拓扑亦是成立的 ($X \times \mathbb{R}$ 中的弱拓扑是 X 的弱拓扑与 $\mathbb{R}$ 的拓扑的乘积). 因为 Banach 空间中的凸集是强闭的当且仅当它是弱闭的, 所以强的 l.s.c. 凸函数是弱 l.s.c. 凸函数. 此种情形, 谈到 l.s.c. 凸函数时, 并不明确指出是何种拓扑. 结合推论 2.29(ii), 可推出下述重要结果.

定理 2.31 设 X 是自反的Banach 空间, $f: X \to \overline{\mathbb{R}}$ 是 l.s.c. 凸函数, 满足至少有一水平集是非空有界的. 则 f 在 X 上的最小值是可取得到的.

Hilbert 空间

设 X 是向量空间. 数乘 (scalar product)$(\cdot,\cdot): X \times X \to \mathbb{R}$ 是对称的正定双线性形式. 即 (i)$x \mapsto (x,y)$ 对所有 $y \in X$ 均是线性的; (ii) 对所有 $x, y \in X, (x,y) = (y,x)$; (iii) 对所有 $x \in X, (x,x) \geqslant 0$ 且 $(x,x) = 0$ 当且仅当 $x = 0$. 若 $(\cdot,\cdot)$ 是数乘, 则实值函数 $\|x\| := (x,x)^{\frac{1}{2}}$ 定义了 X 的范数. 若在此范数下, X 是完备的, 则 X 被称为 Hilbert 空间. (Cauchy-Schwartz) 不等式 $(x,y) \leqslant \|x\|\ \|y\|$ 及平行四边形 (parallelogram) 恒等式

$$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$$

对 Hilbert 空间X 中的任何 x, y 均是成立的.

定义 2.32 称定义在 Hilbert 空间X 上的正常函数f 是强凸的, 若存在 $\alpha > 0$, 对 $x_1, x_2 \in X, t \in [0,1]$, 下述不等式成立

$$f(tx_1 + (1-t)x_2) \leqslant tf(x_1) + (1-t)f(x_2) - \frac{\alpha}{2}t(1-t)\|x_1 - x_2\|^2. \tag{2.19}$$

引理 2.33 设 f 是定义于 Banach 空间X 上的凸的 l.s.c. 函数, 存在 $x_0 \in X$, 满足 $f(x_0) \in \mathbb{R}$. 则

(i) 存在连续的仿射函数 (因而是正常的) 小于 f.

(ii) 若 X 是 Hilbert 空间, f 是强凸的, 则 f 在 X 上于唯一点处取到最小值.

证明 (i) 由推论 2.15, 集合 epi(f) 是凸集, 它是 $X \times \mathbb{R}$ 中包含它的闭半空间的交. 若此族闭半空间中至少有一者是 "非垂直" 的闭半空间, 即存在 $x^* \in X^*$ 及 $\beta \in \mathbb{R}$, 具有形式 $\{(x,\alpha): \alpha \geqslant \langle x^*, x\rangle + \beta\}$(注意到, 由上图的定义, 此不等式不会是相反方向的), 则

$$f(x) \geqslant \langle x^*, x\rangle - \beta, \quad \forall x \in X, \tag{2.20}$$

这即是欲证的结论. 否则, 若所有的闭半空间均是 "竖直" 的闭半空间, 具有形式 $\{(x,\alpha): 0 \geqslant \langle x^*, x\rangle + \beta\}$, 则 $f(x)$ 在其定义域上必等于 $-\infty$, 这与 $f(x_0) \in \mathbb{R}$ 是矛盾的.

(ii) 证 f 的下确界非 $-\infty$. 由强凸性及 (i) 得

$$\begin{aligned}\frac{1}{2}f(x) + \frac{1}{2}f(x_0) &\geqslant f\left(\frac{1}{2}x + \frac{1}{2}x_0\right) + \frac{1}{8}\alpha\|x - x_0\|^2 \\ &\geqslant \frac{1}{2}\langle x^*, x + x_0\rangle - \beta + \frac{1}{8}\alpha\|x - x_0\|^2,\end{aligned}$$

则存在非负常数 $\beta_i \in \mathbb{R}$, 并用 Cauchy-Schwartz 不等式, 得

$$\begin{aligned} f(x) &\geqslant \frac{1}{4}\alpha\|x - x_0\|^2 + \langle x^*, x\rangle - \beta_1 \\ &\geqslant \frac{1}{4}\alpha\|x\|^2 - \beta_2\|x\| - \beta_3 \\ &\geqslant \frac{1}{8}\alpha\|x\|^2 - \beta_4. \end{aligned}$$

这显然表明 $\inf(f) \in \mathbb{R}$.

下证极小化序列 $\{x_k\}$ 是收敛的. 实际上, f 的强凸性可得

$$\begin{aligned} &\frac{\alpha}{8}\|x_n - x_m\|^2 + f\left(\frac{x_n + x_m}{2}\right) - \inf_x f(x) \\ \leqslant &\frac{1}{2}(f(x_n) - \inf_x f(x)) + \frac{1}{2}(f(x_m) - \inf_x f(x)). \end{aligned} \tag{2.21}$$

由上述不等式可得, $\{x_k\}$ 是 Cauchy 序列, 因此收敛到某一 $\bar{x} \in X$. 由 f 的下半连续性,

$$f(\bar{x}) \leqslant \lim f(x_k) = \inf_x f(x),$$

因此 f 于 $\bar{x}$ 处取到最小值. 因为任何极小化序列均是收敛的, f 达到最小的点是唯一的. □

定理 2.34 (Riesz 表示定理) *设 X 是 Hilbert 空间. 则对每一连续的线性泛函 $x^* \in X^*$, 存在向量 $x \in X$, 满足*

$$(x, h) = \langle x^*, h\rangle, \quad \forall h \in X, \tag{2.22}$$

且 $\|x^*\| = \|x\|$.

证明 函数 $f(x) := \frac{1}{2}\|x\|^2 - \langle x^*, x\rangle$ 是强凸的, 因此在唯一点 $\bar{x}$ 处取到最小值, 则对任意 $h \in X$, 有

$$0 \leqslant \lim_{t\downarrow 0} \frac{f(\bar{x} + th) - f(\bar{x})}{t} = (\bar{x}, h) - \langle x^*, h\rangle.$$

对 $-h$ 再写出相应的不等式, 令 $x = \bar{x}$, 即得 (2.22), 且 $\|x^*\| = \|x\|$ 成立. □

由上述定理得 Hilbert 空间 X 的对偶空间 X^* 与 X 是等距等同的, 因此, 每一 Hilbert 空间是自反的.

命题 2.35 *设 X 是 Hilbert 空间, $\{x_n\}$ 是 X 中的序列. 若 $x_n \xrightarrow{w} x$ 且 $\|x_n\| \to \|x\|$, 则 $x_n \to x$.*

证明 显然有

$$\|x_n - x\|^2 = (x_n - x, x_n - x) = \|x_n\|^2 - 2(x_n, x) + \|x\|^2.$$

由 $x_n \xrightarrow{w} x$ 可得 $(x_n, x) \to (x, x)$. 则由 $\|x_n\| \to \|x\|$ 可得 $\|x_n - x\| \to 0$, 证得结论.□

这一结果与 Legendre 形式这一重要概念有关, 见 3.3.2 节. 现在看 Banach 空间的一些例子.

例 2.36 (ℓ_p 空间) 对 $1 \leqslant p < \infty$, 考虑满足 $\sum_{i=1}^{\infty} |x_i|^p < \infty$ 的序列 $x = (x_i)_{i=1}^{\infty}$ 构成的线性空间, 赋予下述范数

$$\|x\|_p := \left(\sum_{i=1}^{\infty} |x_i|^p \right)^{\frac{1}{p}}. \tag{2.23}$$

这是一个可分的 Banach 空间, 记为 ℓ_p. 令 $q \in (1, +\infty)$ 是数 $p \in (1, +\infty)$ 的共轭数, 通过关系 $1/p + 1/q = 1$ 来定义. 则 ℓ_p 的对偶是 ℓ_q. 结果, ℓ_p 等于它的双重对偶, 即对任何 $p \in (1, +\infty)$, ℓ_p 是自反的.

ℓ_1 的对偶是空间 ℓ_∞, 它是有界序列 $(x_i)_{i=1}^{\infty}$ 形成的空间, 赋予范数$\|x\|_\infty := \max\{|x_i| : i = 1, 2, \cdots\}$. 空间 ℓ_1 与 ℓ_∞ 不是自反的, 空间 ℓ_∞ 不是可分的.

例 2.37 ($C(\Omega)$ 空间) 设 Ω 是紧致的 Hausdorff 拓扑空间, 考虑连续函数 $\psi : \Omega \to \mathbb{R}$ 构成的线性空间, 赋予上确界范数 (sup-norm)

$$\|\psi\| := \sup_{\omega \in \Omega} |\psi(\omega)|. \tag{2.24}$$

这是 Banach 空间, 记为 $C(\Omega)$. 由表示定理 (Riesz), $C(\Omega)$ 的对偶空间是由 $(\Omega, \mathcal{B})$ 的 (有限符号) 正则 Borel 测度构成的线性空间, 其中 $\mathcal{B}$ 是 Ω 的 Borel Sigma 代数, 其范数由对应的测度的全变差给出. 即若 μ 是 $(\Omega, \mathcal{B})$ 上的 (有限符号) 测度, $\psi \in C(\Omega)$, 则

$$\langle \mu, \psi \rangle = \int_{\Omega} \psi(\omega) \mathrm{d}\mu(\omega), \tag{2.25}$$

$$\|\mu\| = \sup_{A \in \mathcal{B}} \mu(A) - \inf_{B \in \mathcal{B}} \mu(B). \tag{2.26}$$

除非 Ω 是有限集, $C(\Omega)$ 不是自反的. 若 Ω 是紧致的度量空间, 则 $C(\Omega)$ 是可分的.

Ω 的子集是 Borel 集, 若它可由可数的开集取并、交、补的运算得到. 若 μ 是 Ω 上的 Borel 测度, 则

$$\mu^+(A) := \sup\{\mu(B) : B \subset A, B \in \mathcal{B}\},$$

$$\mu^-(A) := -\inf\{\mu(B) : B \subset A, B \in \mathcal{B}\}$$

被称为 μ 的正与负部分. 注意 μ^+ 与 μ^- 亦是 Ω 上的 Borel 测度, 且对任何 $A \in \mathcal{B}, \mu(A) = \mu^+(A) - \mu^-(A)$. 称 $|\mu|(A) := \mu^+(A) + \mu^-(A)$ 是测度的全变差测度 (total variation measure).

Borel 测度 μ 被称为是非负的, 记为 $\mu \succeq 0$, 若对任何 Borel 集 $A, \mu(A) \geqslant 0$. Borel 测度被称为是正则的 (regular), 若对每一 Borel 集 A, 对任意 $\varepsilon > 0$, 存在闭集 $C \subset \Omega$ 和开集 $V \subset \Omega$, 满足 $C \subset A \subset V$, $|\mu|(A \backslash C) < \varepsilon$, $|\mu|(V \backslash A) < \varepsilon$. 若 Ω 是紧致的度量空间, 则每一 (有限符号的) Ω 上的 Borel 测度是正则的.

设 μ 是 Ω 上的正则 Borel 测度. 考虑 Ω 中满足 $|\mu|(B) = 0$ 的开子集 B 的并集 A. 集合 A 是开的 (因为它是开集的并), 且 $|\mu|(A) = 0$. 实际上, 由于 μ 是正则的, 对任何 $\varepsilon > 0$, 存在闭集 $C \subset A$ 满足 $|\mu|(A \backslash C) < \varepsilon$. 由于 C 是 Ω 中的闭子集, 而 Ω 是紧致的, 有 C 亦是紧致的. 因此, 它可被有限个满足 $\mu(B) = 0$ 的开集 B 覆盖. 因此 $|\mu|(C) = 0$. 从而对任意 $\varepsilon > 0$, $|\mu|(A) \leqslant \varepsilon$, 因此有 $|\mu|(A) = 0$. A 的补集 (即集合 $\Omega \backslash A$) 称为 μ 的支撑(support), 记为 $\operatorname{supp}(\mu)$. 即 $\operatorname{supp}(\mu)$ 是其补集具有零的全变差测度的 Ω 的最小闭子集.

下述例子表明, 类似于推论 2.15 的性质对弱 * 拓扑是不成立的, 即在 Banach 空间的对偶空间中为 (强) 闭凸的集合, 在弱 * 拓扑下不一定是闭的.

例 2.38 考虑集合 (区间) $\Omega := [0,1]$, 令 $X := C[0,1], K$ 是 X^* 的子集. 它由形式为 $\mu = \sum_{n=1}^{\infty} \alpha_n \delta(1/n)$ 的测度组成, 其中 $\delta(t)$ 记在 $t \in [0,1]$ 处的 Dirac 测度, $\sum_{n=1}^{\infty} |\alpha_n| < \infty$. 因此 $\|\mu\| = \sum_{n=1}^{\infty} |\alpha_n|, K$ 与 ℓ_1 是同构的, 因而是闭的. 然而, K 不是弱 * 闭的, 因为在 0 处的 Dirac 测度是 Dirac 测度序列 $\delta(1/n)$ 的弱 * 极限.

例 2.39 ($L_p(\Omega, \mathcal{F}, \mu)$ 空间) 设 $(\Omega, \mathcal{F}, \mu)$ 是测度空间, 即 $\mathcal{F}$ 是 Ω 上的 Sigma 代数, μ 是 $(\Omega, \mathcal{F})$ 上的测度. 对 $1 \leqslant p \leqslant \infty$, 考虑满足

$$\int_{\Omega} |\psi(\omega)|^p \mathrm{d}\mu(\omega) < \infty, \quad \text{若 } p = \infty, \quad \underset{\omega \in \Omega}{\operatorname{supess}} |\psi(\omega)| < \infty$$

的 $\mathcal{F}$ 可测函数 $\psi : \Omega \to \mathbb{R}$ 等价类构成的线性空间, 赋予范数

$$\|\psi\|_p = \left(\int_{\Omega} |\psi(\omega)|^p \mathrm{d}\mu(\omega) \right)^{\frac{1}{p}}, \quad \text{若 } p < \infty, \tag{2.27}$$

$$\|\psi\|_{\infty} = \underset{\omega \in \Omega}{\operatorname{supess}} |\psi(\omega)|, \quad \text{若 } p = \infty, \tag{2.28}$$

这里, 等价类由只在 μ 测度为零的集合上取不同值的函数构成, 这是 Banach 空间, 记为 $L_p(\Omega, \mathcal{F}, \mu)$. 尤其, 若 Ω 是 $\mathbb{R}^n$ 的子集, μ 是 Lebesgue 测度, 则记为 $L_p(\Omega)$. 对于 $p \in (0,1), L_p(\Omega, \mathcal{F}, \mu)$ 是自反的 Banach 空间, 其对偶空间是 $L_q(\Omega, \mathcal{F}, \mu)$, 其中, $q \in (1, \infty)$ 同样是满足 $1/p + 1/q = 1$ 的共轭数. $L_1(\Omega, \mathcal{F}, \mu)$ 的对偶空间是 $L_{\infty}(\Omega, \mathcal{F}, \mu)$, 但相反之结论不真, 除非 $L_1(\Omega, \mathcal{F}, \mu)$ 是有限维的. 因此, 通常来讲, $L_1(\Omega, \mathcal{F}, \mu)$ 与 $L_{\infty}(\Omega, \mathcal{F}, \mu)$ 不是自反的.

2.1.4 锥、对偶性与回收锥

这一节设 X 与 X^* 是 (局部凸的拓扑) 成对空间 (见定义 2.26). 注意, X 与 X^* 赋予相容的拓扑时, X 与 X^* 间存在完整的对称, 关于对偶性的所有结论在两个方向上均有效. 尤其当 X 是非自反的 Banach 空间, X^* 是它的拓扑对偶, 赋予弱 * 拓扑, 则在公式 (2.32) 与 (2.33) 中的闭包运算应理解为弱 * 拓扑 (而不是强拓扑).

X 中的非空子集 C 被称为是锥 (cone), 若对任何 $x \in C$, 对任意的 $t \geqslant 0$, 有 $tx \in C$. 对锥 $C \subset X$, 它的极 (polar) (或其负对偶) 锥 C^- 定义为

$$C^- := \{x^* \in X^* : \langle x^*, x\rangle \leqslant 0, 对所有 x \in C\}. \tag{2.29}$$

还可以定义双极 (bipolar) 锥

$$(C^-)^- := \{x \in X : \langle x^*, x\rangle \leqslant 0, 对所有 x^* \in C^-\}. \tag{2.30}$$

由上述定义知, C^- 可表示为闭 (且凸的) 的半空间的交, 因此, 它在 X^* 的相容拓扑下是凸的闭的. 类似地, $(C^-)^-$ 在 X 的相容拓扑下是凸的闭的.

命题 2.40 *设 $C \subset X$ 是凸锥. 则 $(C^-)^- = \mathrm{cl}(C)$. 尤其, 若 C 是闭的, 则 $(C^-)^- = C$.*

证明 由定义 (2.29) 与 (2.30) 得 $C \subset (C^-)^-$. 由于 $(C^-)^-$ 是闭的, 这表明 $\mathrm{cl}(C) \subset (C^-)^-$. 相反地, 假设 $x_0 \notin \mathrm{cl}(C)$. 由第二分离定理2.14, 存在非零 $x^* \in X^*$ 强分离 x_0 与 $\mathrm{cl}(C)$. 因为 C 是锥, 得到对所有 $x \in \mathrm{cl}(C)$, $\langle x^*, x\rangle \leqslant 0$ 且 $\langle x^*, x_0\rangle > 0$. 结果得到 $x^* \in C^-$ 且 $x_0 \notin (C^-)^-$, 因此 $(C^-)^- \subset \mathrm{cl}(C)$, 这就证得结论. □

若 C 是线性空间, 则 C^- 即 C 的正交补空间

$$C^\perp := \{x^* \in X^* : \langle x^*, x\rangle = 0, \forall x \in C\}.$$

尤其, 若 $x^* \in X^*$, 则 $[\![x^*]\!]^- = [\![x^*]\!]^\perp = \ker x^*$, 其中 $\ker x^*$ 是 x^* 的零空间. $[\![x^*]\!]$ 记由 x^* 生成的空间 (若 $x^* \neq 0$, 是一维的). 可直接验证, 若 C_1 与 C_2 是 X 或 X^* 的两个锥, 则

$$(C_1 + C_2)^- = C_1^- \cap C_2^-. \tag{2.31}$$

由 (2.31) 可得, $C_1^- \cap C_2^-$ 的极锥与 $(C_1 + C_2)^-$ 的极锥相同. 结果, 若锥 C_1 与 C_2 是凸的, 则 $C_1^- \cap C_2^-$ 的极锥为锥 $C_1 + C_2$ 的拓扑包. 记 $K_1 = C_1^-, K_2 = C_2^-$. 有 $(K_1 \cap K_2)^-$ 与锥 $K_1^- + K_2^-$ 的拓扑包相同. 因为任何闭凸锥均可表示为极锥, 可得, 若 K_1 与 K_2 是 X(或 X^*) 中的两个闭凸锥, 则

$$(K_1 \cap K_2)^- = \mathrm{cl}\{K_1^- + K_2^-\}. \tag{2.32}$$

若 K_3 是另一闭凸锥, 则

$$\begin{aligned}(K_1 \cap K_2 \cap K_3)^- &= (K_1 \cap (K_2 \cap K_3))^- \\ &= \mathrm{cl}\{K_1^- + (K_2 \cap K_3)^-\} \\ &= \mathrm{cl}\{K_1^- + \mathrm{cl}\{K_2^- + K_3^-\}\} \\ &= \mathrm{cl}\{K_1^- + K_2^- + K_3^-\}.\end{aligned}$$

由归纳法可得, 若 $K_1, \cdots, K_p$ 是闭凸锥, 则

$$\left(\bigcap_{i=1}^{p} K_i\right)^- = \mathrm{cl}\left\{\sum_{i=1}^{p} K_i^-\right\}. \tag{2.33}$$

命题 2.41 设 X 是局部凸的拓扑向量空间, 令

$$C := \left\{\sum_{i=1}^{n} \lambda_i a_i : \lambda_i \geqslant 0, i = 1, \cdots, n\right\}$$

为由有限个向量 $a_i \in X, i = 1, \cdots, n$ 生成的锥. 则 C 是闭的且存在 $c > 0$, 满足对每一 $y \in C$, 存在 $\lambda \in \mathbb{R}_+^n$, 使

$$\|\lambda\| \leqslant c\|y\| \quad 且 \quad y = \sum_{i=1}^{n} \lambda_i a_i. \tag{2.34}$$

证明 对 $y \in C$, 考虑

$$M(y) := \left\{\lambda \in \mathbb{R}_+^n : y = \sum_{i=1}^{n} \lambda_i a_i\right\}.$$

令

$$\mathrm{supp}(\lambda) := \{i : \lambda_i \neq 0, i = 1, \cdots, n\}$$

是 λ 的支撑. 称 $\lambda \in M(y)$ 具有最小的支撑, 若不存在另外的 $\lambda' \in M(y)$, 它的支撑严格包含在 λ 的支撑里. 令 $\lambda \in M(y)$ 具有最小的支撑, 置 $I^* := \mathrm{supp}(\lambda)$. 下证存在不依赖于 μ 的常数 $c_{I^*} > 0$, 有 $\left\|\sum_{i \in I^*} \mu_i a_i\right\| \geqslant c_{I^*}\|\mu\|$. 否则, 存在序列 $\{\mu^n\}$ 满足 $\left\|\sum_{i \in I^*} \mu_i^n a_i\right\| \leqslant n^{-1}\|\mu^n\|$. 则 $\mu^n/\|\mu^n\|$ 有极限点 μ 满足 $\|\mu\| = 1$ 且 $\left\|\sum_{i \in I^*} \mu_i a_i\right\| = 0$. 若有必要可将 μ 换成 $-\mu$, 不妨设 $\min_i \mu_i < 0$, 因此, 存在 $t > 0$, 满足 $\lambda + t\mu \geqslant 0$, 而 $\min_i(\lambda_i + t\mu_i) = 0$. 显然, 这与 λ 具有最小支撑相矛盾. 可见论断为真. 由于只有有限个可能的 I^*, 令 $c = \min\{c_{I^*}\}$, 即得 (2.34).

下证 C 的闭性. 因为 C 包含在由 $a_i \in X, i=1,\cdots,n$ 生成的线性空间 Z 中, 由命题 2.18 得后者是闭的, 只需要证明 C 是 Z 的闭子集. 因为由 Z 引导的拓扑与有限维向量空间的标准拓扑是同构的, 只需要证: 若 $\{y_i\}_{i\in\mathbb{N}} \subset C, y_i \to y$, 则 $y \in C$. 因为由 (2.33) 可知, 存在有界序列 $\lambda^i \in M(y^i)$, 对该序列中的任一极限点λ, 可有 $\lambda \geqslant 0$ 且 $\sum_{i=1}^{n} \lambda_i a_i = y$, 因此 $\lambda \in M(y)$. □

命题 2.42 *设 $a_i \in X^*, i=1,\cdots,p$, 考虑锥*

$$K := \{x \in X : \langle a_i, x\rangle \leqslant 0, i=1,\cdots,p\}.$$

则其极锥可表示为下述形式

$$K^- = \left\{\sum_{i=1}^{p} \lambda_i a_i : \lambda_i \geqslant 0, i=1,\cdots,p\right\}. \tag{2.35}$$

证明 因为对给定 $i \in \{1,\cdots,p\}$, 集合 $\{x \in X : \langle a_i, x\rangle \leqslant 0\}$ 是闭凸锥 $\{ta_i : t \geqslant 0\}$ 的极锥, 有 $\{x \in X : \langle a_i, x\rangle \leqslant 0\}$ 的极锥是 $\{ta_i : t \geqslant 0\}$. 由 (2.33) 得 K^- 是 (2.35) 的右端给出的集合的闭包. 注意到, 由命题 2.41 知 (2.35) 的右端的集合是闭的, 因此证得结论. □

上述结果的推广将在命题 2.201 中给出.

回收锥

考虑凸集 $S \subset X$. 回收锥(recession cone) S^∞ 定义为

$$S^\infty := \{x \in X : x + S \subset S\}. \tag{2.36}$$

从上述定义容易得到, S^∞ 是凸的, 且若 $x \in S^\infty$, 则对任何整数 $n > 0$ 有 $nx \in S^\infty$. 因此, 由 S^∞ 的凸性得 S^∞ 是锥. 不难看到, 若 S 是闭集, 则 S^∞ 亦是闭的. 等价地, 回收锥S^∞ 定义为对任何 $\bar{x} \in S$ 及任何 $t \geqslant 0$, 均有 $\bar{x} + tx \in S$ 的 $x \in X$ 的集合. 进一步, 若集合 S(因而 S^∞) 是闭的, 则 S^∞ 由满足对某一 $\bar{x} \in S$ 及对所有 $t \geqslant 0$, 均有 $\bar{x} + tx \in S$ 的所有 $x \in X$ 组成.

很显然, 若凸集 S 的回收锥含有非零向量, 则 S 是无界的. 进一步, 若空间 X 是有限维的, 则非空凸集 $S \subset X$ 是有界的当且仅当 $S^\infty = \{0\}$. 下面的例子表明, 这一点在无穷维空间中是不成立的.

例 2.43 考虑 Hilbert 空间$X := \ell_2$ 及闭凸子集

$$S := \{(x_i) \in \ell_2 : |x_i| \leqslant 1, i=1,2,\cdots\}.$$

若 $x_n = (x_i^n)$ 满足对 $1 \leqslant i \leqslant n$ 有 $x_i^n = 1, x_i^n = 0, i > n$, 则 $x_n \in S$. 显然, $\|x_n\| = \sqrt{n}$, 因此 S 是无界的. 另一方面, 不难验证 $S^\infty = \{0\}$.

2.2　方向可微性与切锥

2.2.1　一阶方向导数

令 X 与 Y 是赋范向量 (线性) 空间, 考虑映射 $g : X \to Y$.

定义 2.44　称 g 在 $x \in X$ 处沿方向 $h \in X$ 是方向可微的, 若极限

$$g'(x, h) := \lim_{t \downarrow 0} \frac{g(x + th) - g(x)}{t} \tag{2.37}$$

存在. 若 g 在 x 处沿每一方向 $h \in X$ 均是方向可微的, 则称 g 在 x 处是方向可微的.

不难验证, 方向导数$g'(x, h)$, 若它存在, 关于 h 是正齐次的, 即对任何 $t \geqslant 0$, $g'(x, th) = tg'(x, h)$. 称 G 在 x 处是 Gâteaux 可微的 (Gâteaux differentiable), 若 g 在 x 处是方向可微的且方向导数$g'(x, h)$ 关于 h 是线性的连续的. 即 $g'(x, \cdot) : X \to Y$ 是连续的线性算子. 我们记这一算子 (当它存在时) 为 $Dg(x)$, 即 $Dg(x)h = g'(x, h)$.

一个更强的可微性是 Hadamard 意义下的可微性.

定义 2.45　称 g 在 x 处是 Hadamard 意义下方向可微的, 若对所有的 h, 方向导数$g'(x, h)$ 存在, 并且

$$g'(x, h) = \lim_{\substack{t \downarrow 0 \\ h' \to h}} \frac{g(x + th') - g(x)}{t}. \tag{2.38}$$

若 $g'(x, h)$ 关于 h 还是线性的, 则称 g 在 x 处是 Hadamard 可微的 (Hadamard differentiable).

条件 (2.38) 可以等价地表述为下述形式, 即对任何 $h_n \to h, t_n \downarrow 0$, 有

$$g'(x, h) = \lim_{n \to \infty} \frac{g(x + t_n h_n) - g(x)}{t_n}.$$

命题 2.46　若 g 在 x 处是 Hadamard 意义下方向可微的, 则方向导数$g'(x, \cdot)$ 在 X 上是连续的.

证明　设 g 在 x 处是 Hadamard 意义下方向可微的, 考虑两个方向 $h, h^* \in X$. 不妨设 $g(x) = 0$. 则对任意的 $\varepsilon > 0$ 及充分小的 $t > 0$, 有

$$\|g'(x, h^*) - t^{-1} g(x + th^*)\| \leqslant \varepsilon.$$

由 (2.38), 只要 h^* 充分接近 $h, t > 0$ 充分小, 就有

$$\|g'(x, h) - t^{-1} g(x + th^*)\| \leqslant \varepsilon.$$

结合这两个不等式得

$$\|g'(x,h)-g'(x,h^*)\| \leqslant 2\varepsilon.$$

由 ε 的任意性, 结论得证. □

上述命题表明, 若 g 在 x 处是 Hadamard 意义下方向可微的, 且 $g'(x,\cdot)$ 是线性的(即 g 在 x 处是 Hadamard 可微的), 则线性映射 $g'(x,\cdot): X \to Y$ 自动是连续的. Hadamard 意义下的方向可微性的另一重要性质是它使得链式法则成立. 设 X, Y 与 Z 是线性赋范空间, $g: X\to Y, f: Y\to Z$, 考虑复合映射 $f\circ g: X\to Z$.

命题 2.47 设 g 在 x 处是方向可微的, f 于 $y=g(x)$ 处是 Hadamard 方向可微的. 则复合映射 $f\circ g$ 在 x 处是方向可微的, 且下述链式法则成立

$$(f\circ g)'(x,h)=f'(y,g'(x,h)). \tag{2.39}$$

证明 因为 g 在 x 处是方向可微的, 对 $t\geqslant 0$, 有

$$g(x+th)=g(x)+tw+o(t),$$

其中 $w=g'(x,h)$. 进一步, 因为 f 在 $y=g(x)$ 处是 Hadamard 方向可微的, 有

$$f(g(x+th))=f(g(x)+tw+o(t))=f(y)+tf'(g,w)+o(t).$$

这表明复合映射在 x 处是方向可微的且方向导数满足链式法则 (2.39). □

不难验证, 在上述命题中, 若映射 g 在 x 处是 Hadamard 方向可微的, 则复合映射也是 Hadamard 方向可微的.

另一个经常用到的可微性的概念是 Fréchet 意义的可微性.

定义 2.48 称 g 在 x 处是 Fréchet 意义下方向可微的, 若 g 在 x 处是方向可微的且

$$g(x+h)=g(x)+g'(x,h)+o(\|h\|), \quad h\in X. \tag{2.40}$$

若 $g'(x,\cdot)$ 是线性的连续的, 称 g 在 x 处是 Fréchet 可微的 (Fréchet differentiable). 注意到 (2.37) 与 (2.40) 都不能推出 $g'(x,\cdot)$ 的连续性, 如正齐次非连续函数.

现在设 g 在开集 $S\subset X$ 上是 Gâteaux 可微的, 其相应的导数 $Dg(x)$ 在 S 上是连续的 (在 $\mathcal{L}(X,Y)$ 的算子范数拓扑下, 即关于范数 $\|A\|=\sup\limits_{x\in B_X}\|Ax\|$, $A\in\mathcal{L}(X,Y)$). 此时, 称 g 在 S 上是连续可微的 (continuously differentiable). 令 $[x,x+h]\subset S$, 考虑 $\phi(t):=g(x+th)$. 则 $\phi'(t)=Dg(x+th)h$, 且因为 $Dg(\cdot)$[①]是连续的, 则

$$\phi(1)-\phi(0)=\int_0^1\phi'(t)\mathrm{d}t.$$

① 原著中是 $DG(\cdot)$.

得到下述积分表示 (中值定理)

$$g(x+h)-g(x)=\int_0^1 Dg(x+th)h\mathrm{d}t. \tag{2.41}$$

由 (2.41) 得到

$$\|g(x+h)-g(x)-Dg(x)h\|\leqslant\|h\|\int_0^1\|Dg(x+th)-Dg(x)\|\mathrm{d}t,$$

因此, 由 $Dg(\cdot)$ 的连续性得

$$g(x+h)=g(x)+Dg(x)h+o(\|h\|), \tag{2.42}$$

即 g 在 x 处是 Fréchet 可微的. 因此, 对连续可微的映射而言, Gâteaux 与 Fréchet 导数是等价的.

在上述方向可微性概念间存在各种关系. 映射 $g: X\to Y$ 称为是模为 c 的在集合 $S\subset X$ 上为 Lipschitz 连续的, 若对于 $x_1, x_2\in S$, 满足

$$\|g(x_1)-g(x_2)\|\leqslant c\|x_1-x_2\|. \tag{2.43}$$

命题 2.49　*设 g 在 x 处是方向可微的, 且在 x 的邻域上是以 c 为模的 Lipschitz 连续映射. 则 g 在 x 处是 Hadamard 意义下方向可微的, 且方向导数$g'(x,\cdot)$ 在 X 上是 Lipschitz 连续的, 模为 c.*

证明　因为有

$$\|g'(x,h_1)-g'(x,h_2)\|\leqslant\lim_{t\downarrow 0}\frac{\|g(x,th_1)-g(x,th_2)\|}{t}\leqslant c\|h_1-h_2\|$$

(注意, 上述极限存在), 因此 $g'(x,\cdot)$ 在 X 上是模为 Lipschitz 连续的 c. 还有

$$\frac{1}{t}[g(x+th^*)-g(x)]=\frac{1}{t}[g(x+th)-g(x)]+\frac{1}{t}[g(x+th^*)-g(x+th)],$$

由 Lipschitz 连续性有

$$t^{-1}\|g(x+th^*)-g(x+th)\|\leqslant c\|h-h^*\|$$

对 t 充分小成立, 因此由 (2.37) 可得 (2.38). □

若空间 X 是有限维的, 则由 Hadamard 方向可微性可推出 Fréchet 方向可微性. 若 g 是 Fréchet 方向可微的且 $g'(x,\cdot)$ 是连续的, 则它亦是 Hadamard 方向可微的. 进一步, 若 X 是有限维的, g 是局部 Lipschitz 连续的, 则上述方向可微性概念是等价的.

下述例子表明, 可能发生凸的连续函数是 Gâteaux(Hadamard) 可微的但不是 Fréchet 可微的情形.

例 2.50 考虑 Hilbert 空间$X := L_2[0,1]$, 集合 $K \subset L_2[0,1]$ 由几乎处处非负值函数 $x(\cdot) \in L_2[0,1]$ 构成, 函数 $f(x) := \text{dist}(x, K) : X \to \mathbb{R}$. 函数 $f(x)$ 是连续函数, 因集 K 是凸的, 它还是凸函数. 不难得到 $f(x) = \|x_-\|$, 其中 $x_-(t) := \min\{x(t), 0\}$, $t \in [0,1]$. 考虑函数 $x_0(t) = 1$, $t \in [0,1]$. 显然 $x_0 \in K$, 不难检验, f 在 x_0 处 Gâteaux 可微的, $Df(x_0) = 0$. 另一方面, 考虑函数序列 $h_n(t) := -(1+2\alpha_n)t^{\alpha_n}$, $t \in (0,1]$, 其中 α_n 是满足 $\alpha_n \downarrow (-1/2)$ 的数列. 注意到 $\|h_n\| = (1+2\alpha_n)^{1/2} \to 0$. 直接计算可验证 $f(x_0 + h_n)/\|h_n\| \to e^{-1} \neq 0$. 结果得到函数 f 于 x_0 处不是 Fréchet 可微的.

2.2.2 二阶导数

现在讨论映射 $g : X \to Y$ 的二阶导数. 设映射 g 在点 x 的一邻域内是 Gâteaux 可微的, 考虑映射 $Dg(\cdot) : X \to \mathcal{L}(X, Y)$, 其中 $\mathcal{L}(X, Y)$ 视为赋予算子范数下的赋范空间. 这一映射在 x 处的 (Gâteaux, Hadamard 或 Fréchet) 导数视为 g 在 x 处的二阶导数.

从一个不同的角度 (然而是等价的角度) 来研究二阶导数是合适的. 映射 $B : X \times X \to Y$ 称为是双线性的(bilinear), 如果对任何 $x \in X$, $B(x, \cdot)$ 与 $B(\cdot, x)$ 在 X 上均是线性的. 双线性映射B 是连续的当且仅当它是有界的, 即范数

$$\|B\| := \sup\{\|B(x_1, x_2)\| : \|x_1\| \leqslant 1,\ \|x_2\| \leqslant 1\} \tag{2.44}$$

是有限的. 双线性映射B 被称为是对称的 (symmetric), 若对任何 $x_1, x_2 \in X$, 有 $B(x_1, x_2) = B(x_2, x_1)$. 注意到, 任何双线性映射$B(x_1, x_2)$, 通过下述方式都联系着对称的双线性映射

$$\overline{B}(x_1, x_2) := \frac{1}{2}[B(x_1, x_2) + B(x_2, x_1)],$$

显然有 $\overline{B}(x, x) = B(x, x)$.

考虑线性连续映射$\mathcal{A} : X \to \mathcal{L}(X, Y)$. 将这一映射同下述的双线性映射相联系

$$B(x_1, x_2) := (\mathcal{A}x_1)x_2. \tag{2.45}$$

直接验证知, 上述定义的映射实际上是双线性的. 进一步

$$\|B(x_1, x_2)\| \leqslant \|\mathcal{A}x_1\|\ \|x_2\| \leqslant \|\mathcal{A}\|\ \|x_1\|\ \|x_2\|,$$

因此有 $\|B\| \leqslant \|\mathcal{A}\|$. 相反地, 若 $B(x_1, x_2)$ 是一双线性映射, 将公式 (2.45) 以相反的顺序写出来就得到线性映射 $\mathcal{A} : X \to \mathcal{L}(X, Y)$, 即

$$(\mathcal{A}x_1)(\cdot) := B(x_1, \cdot).$$

显然, 有 $\|\mathcal{A}\| \leqslant \|B\|$. 因此, 联系 $\mathcal{A}$ 与 B 的映射是等距的且映上的. 这使得我们把空间 $\mathcal{L}(X, \mathcal{L}(X, Y))$ 与连续的双线性映射构成的 Banach 空间等同起来.

设映射 g 在某点 x 的邻域内是 Gâteaux 可微的. 则在 x 处的二阶 Gâteaux 导数 $D^2g(x) \in \mathcal{L}(X, \mathcal{L}(X, Y))$ 定义为

$$D^2g(x)h := \lim_{t \downarrow 0} \frac{Dg(x+th) - Dg(x)}{t}, \tag{2.46}$$

其中上述极限取为关于 $\mathcal{L}(X, Y)$ 的算子范数拓扑. 称 g 在 x 处是二次 Gâteaux 可微的, 若上述极限对所有 $h \in X$ 均存在且 $D^2g(x)h$ 是 h 的线性的连续的, 即 $D^2g(x) \in \mathcal{L}(X, \mathcal{L}(X, Y))$. 称 g 在 x 处是二次连续可微的 (twice continuously differentiable), 若在 x 的邻域内它是二次 Gâteaux 可微的且映射 $Dg : X \to Y$ 与 $D^2g : X \to \mathcal{L}(X, Y)$ 在 x 的邻域内是连续的.

引理 2.51　设 X 与 Y 是 Banach 空间, $g : X \to Y$ 在点 x 处是二阶连续可微的. 则对 X 中的所有 h_1 与 h_2, 有

$$[D^2g(x)h_2]h_1 = [D^2g(x)h_1]h_2. \tag{2.47}$$

证明　观察到, 对任意 $h_1, h_2 \in X$ 及 $t \in \mathbb{R}$, 有

$$[D^2g(x)th_1]th_2 = t^2[D^2g(x)h_1]h_2,$$

只需证 (2.47) 对 $0 \in X$ 的邻域中的 h_1 与 h_2 成立即可. 考虑

$$\gamma(h_1, h_2) := g(x + h_1 + h_2) - g(x + h_1) - g(x + h_2) + g(x).$$

由中值定理(见 (2.41)), 对零的邻域中的 h_1, h_2, 有

$$\begin{aligned}\gamma(h_1, h_2) &= \int_0^1 Dg(x + h_1 + th_2)h_2 \mathrm{d}t - \int_0^1 Dg(x + th_2)h_2 \mathrm{d}t \\ &= \int_0^1 \int_0^1 [D^2g(x + sh_1 + th_2)h_1]h_2 \mathrm{d}s\mathrm{d}t.\end{aligned} \tag{2.48}$$

结果, 对任意 $\varepsilon > 0$, 对充分接近于零的 h_1, h_2, 有

$$\gamma(h_1, h_2) = [D^2g(x)h_1]h_2 + r(h_1, h_2), \tag{2.49}$$

其中 $\|r(h_1, h_2)\| \leqslant \varepsilon\|h_1\| \, \|h_2\|$. 同样可得到

$$\gamma(h_1, h_2) = [D^2g(x)h_2]h_1 + r'(h_1, h_2), \tag{2.50}$$

其中 $\|r'(h_1,h_2)\| \leqslant \varepsilon\|h_1\|\ \|h_2\|$ 当 h_1, h_2 充分接近于 0 时成立. 由于 $\gamma(h_1,h_2)$ 是 h_1 与 h_2 的对称函数, 对任意 $\varepsilon > 0$, 对充分接近于零的 h_1, h_2, 有

$$\|[D^2g(x)h_2]h_1 - [D^2g(x)h_1]h_2\| \leqslant \varepsilon\|h_1\|\ \|h_2\|. \tag{2.51}$$

将 h_1 替换为 th_1, h_2 替换为 th_2, $t \in \mathbb{R}$, 可得到对 $t > 0$ 充分小, (2.51) 的左端会任意小, 因此 (2.47) 成立. □

上述引理意味着对应 $D^2g(x)$ 的双线性映射是对称的. 视 $D^2g(x)$ 为 (对称的) 双线性映射.

设 $g(\cdot)$ 是二次连续可向映射. 则 $\phi(t) := g(x+th)$ 也是二次连续可微的, 因此可以得到

$$\phi(1) = \phi(0) + \phi'(0) + \int_0^1 \phi''(t)(1-t)\mathrm{d}t.$$

进一步, $\phi'(0) = Dg(x)h$, $\phi''(t) = D^2g(x+th)(h,h)$. 因此得到下述展式

$$g(x+h) = g(x) + Dg(x)h + \int_0^1 D^2g(x+th)(h,h)(1-t)\mathrm{d}t. \tag{2.52}$$

由 D^2g 的连续性可得

$$g(x+h) = g(x) + Dg(x)h + \frac{1}{2}D^2g(x)(h,h) + o(\|h\|^2). \tag{2.53}$$

公式 (2.52) 与 (2.53) 是 $g(\cdot)$ 在 x 处的二阶 Taylor 展式, (2.52) 的余项是积分形式.

后面的讨论主要用二阶 Taylor 展式 (2.53), 其中 $D^2g(x)(\cdot,\cdot)$ 是对称的连续的双线性映射.

2.2.3 增广实值函数的方向上图导数

设 X 是局部凸的拓扑向量空间. 考虑增广实值函数 $f: X \to \overline{\mathbb{R}}$, 一点 $x \in X$, 满足 $f(x)$ 有限. f 在 x 处的上、下方向导数分别定义为

$$f'_+(x,h) := \limsup_{t\downarrow 0} \frac{f(x+th)-f(x)}{t}, \tag{2.54}$$

$$f'_-(x,h) := \liminf_{t\downarrow 0} \frac{f(x+th)-f(x)}{t}. \tag{2.55}$$

称 f 在 x 处沿方向 h 是方向可微的, 若 $f'_+(x,h) = f'_-(x,h)$.

应该指出, 本节中 f 的方向可微性与 2.2.1 节的方向可微性稍有不同. 这里, 方向导数 $f'(x,\cdot)$ 是增广实值函数, 可取值 $-\infty$ 或 $+\infty$. 当然, 若存在 $h \in X$, $f'(x,h)$ 有限, 则方向导数与 2.2.1 节中相应的方向导数是相同的.

引用增广实值函数的方向上图导数是有用的. 为此, 先回顾上图收敛分析 (epi-convergence analysis) 中的某些基本定义和结果.

对 X 中的子集类 A_t, 其中 t 是参数, 它可以是实值的, 或更一般地, 是度量空间的元素, 回顾 Painlevé Kuratowski 意义下的上、下极限.

定义 2.52 设 X 是 Banach 空间. 下述集合被称为 X 中的子集参数类 A_t 的上极限与下极限

$$\limsup_{t\to t_0} A_t := \{x \in X : \liminf_{t\to t_0}[\operatorname{dist}(x, A_t)] = 0\}, \tag{2.56}$$

$$\liminf_{t\to t_0} A_t := \{x \in X : \limsup_{t\to t_0}[\operatorname{dist}(x, A_t)] = 0\}. \tag{2.57}$$

由定义易知, 上极限与下极限均是闭集. 这些集合也可以通过序列的形式如下描述. 上极限 $\limsup\limits_{t\to t_0} A_t$ 可定义为这样的 x 的集合：存在序列 $t_n \to t_0$, 存在 $x_n \in A_{t_n}$, 满足 $x_n \to x$. 类似地, 下极限 $\liminf\limits_{t\to t_0} A_t$ 可定义为这样的 x 的集合：对每一序列 $t_n \to t_0$, 可找到 $x_n \in A_{t_n}$, 满足 $x_n \to x$. 由上述定义很容易得

$$\liminf_{t\to t_0} A_t \subset \limsup_{t\to t_0} A_t.$$

若上式中等式成立, 称 A_t 在 t_0 处有极限.

上、下集合极限还可以表示为

$$\limsup_{t\to t_0} A_t := \bigcap_{\varepsilon>0} \operatorname{cl}\Bigg(\bigcup_{\substack{|t-t_0|<\varepsilon \\ t\neq t_0}} A_t\Bigg), \tag{2.58}$$

$$\liminf_{t\to t_0} A_t := \bigcap_{\substack{t_n\to t_0 \\ t_n\neq t_0}} \bigcap_{n} \operatorname{cl}\Bigg(\bigcup_{i\geqslant n} A_{t_i}\Bigg). \tag{2.59}$$

上述表示式不涉及范数或距离, 可用于定义一般拓扑向量空间中子集参数类 A_t 的上极限与下极限.

设 $f_t : X \to \overline{\mathbb{R}}$ 是一类增广实值函数. f_t 当 $t \to t_0$ 时的下、上上图极限(lower and upper epilimits) 被定义为上图分别是 f_t 的上图的集合极限 $\limsup\limits_{t\to t_0}(\operatorname{epi} f_t)$ 与 $\liminf\limits_{t\to t_0}(\operatorname{epi} f_t)$ 的函数, 这一定义是有意义的. 事实上, 集合 $S \subset X \times \mathbb{R}$ 是增广实值函数的上图的充分必要条件是它满足下述两个性质：

(i) 若 $(x, t) \in S, t' > t$, 则 $(x, t') \in S$.

(ii) 若存在 $x \in X$, $t^* = \inf\{t : (x, t) \in S\}$, 则 $(x, t^*) \in S$.

不难看到, 这些性质在下集合收敛与上集合收敛下是被保持的.

注意, 由于下、上集合极限是闭集, 下、上上图极限函数具有闭的上图, 因此它们是下半连续的.

同样地, 下上图极限与上上图极限可表述为

$$\mathrm{e}-\liminf_{t\to t_0} f_t(x)=\sup_{V\in\mathcal{V}(x)}\liminf_{t\to t_0}\inf_{x'\in V} f_t(x'),\tag{2.60}$$

$$\mathrm{e}-\limsup_{t\to t_0} f_t(x)=\sup_{V\in\mathcal{V}(x)}\limsup_{t\to t_0}\inf_{x'\in V} f_t(x'),\tag{2.61}$$

其中 $\mathcal{V}(x)$ 是 x 的邻域的基本系统 (基), 或

$$\mathrm{e}-\liminf_{t\to t_0} f_t(x)=\liminf_{\substack{t\to t_0\\ x'\to x}} f_t(x'),\tag{2.62}$$

$$\mathrm{e}-\limsup_{t\to t_0} f_t(x)=\sup_{\{t_n\}\in\Sigma_0}\left(\liminf_{\substack{n\to\infty\\ x'\to x}} f_{t_n}(x')\right),\tag{2.63}$$

其中 Σ_0 记所有满足收敛到 t_0 且 $\forall n$ 均有 $t_n\neq t_0$ 的序列 $\{t_n\}$ 的集合, 这里, “$x'\to x$”时, 允许 x' 等于 x (在定义上图导数时也采用类似的约定).

增广实值函数的序列 $\{f_n\}$ 的下上图极限, 上上图极限可以类似地定义. 下上图极限总是不超过上上图极限的. 序列 $\{f_n\}$ 称为上图收敛(epiconverge) 到函数 f, 记为 $f_n\xrightarrow{e} f$, 若 $\{f_n\}$ 的下上图极限与上上图极限相等, 它们等于 f. 若 X 是一 Banach 空间, 可以证明, $f_n\xrightarrow{e} f$ 当且仅当对所有的 $x\in X$, 下述两个条件成立:

(i) 对任何收敛到 x 的序列 $\{x_n\}$, 有

$$\liminf_{n\to\infty} f_n(x_n)\geqslant f(x).\tag{2.64}$$

(ii) 存在收敛到 x 的序列 $\{x_n\}$ 满足

$$\limsup_{n\to\infty} f_n(x_n)\leqslant f(x).\tag{2.65}$$

由上面的讨论可知, 上图极限函数是下半连续的.

设 $f:X\to\overline{\mathbb{R}}$ 是增广实值函数, 在 $x\in X$ 处 $f(x)$ 有限, f 在 x 处的下方向上图导数与上方向上图导数分别定义为

$$f_-^{\downarrow}(x,\cdot):=\mathrm{e}-\liminf_{t\downarrow 0}\frac{f(x+t\cdot)-f(x)}{t},\tag{2.66}$$

$$f_+^{\downarrow}(x,\cdot):=\mathrm{e}-\limsup_{t\downarrow 0}\frac{f(x+t\cdot)-f(x)}{t},\tag{2.67}$$

其中 “$\cdot$” 记上图极限所计算的变量. 由 (2.62) 与 (2.63), 可将这些导数表示为下述的等价形式

$$f_-^{\downarrow}(x,h)=\liminf_{\substack{t\downarrow 0\\ h'\to h}}\frac{f(x+th')-f(x)}{t},\tag{2.68}$$

$$f_+^{\downarrow}(x,h) = \sup_{\{t_n\}\in\Sigma} \left(\liminf_{\substack{n\to\infty \\ h'\to h}} \frac{f(x+t_nh')-f(x)}{t_n} \right), \tag{2.69}$$

其中 Σ 记收敛到 0 的正实数序列 $\{t_n\}$ 的集合. 由于上图极限函数是下半连续的, 有 $f_-^{\downarrow}(x,\cdot)$ 与 $f_+^{\downarrow}(x,\cdot)$ 是下半连续的正齐次函数. 还有

$$f_-^{\downarrow}(x,h) \leqslant f_+^{\downarrow}(x,h), \quad f_-^{\downarrow}(x,h) \leqslant f'_-(x,h), \quad f_+^{\downarrow}(x,h) \leqslant f'_+(x,h). \tag{2.70}$$

注意到, 若 X 是 Banach 空间, f 在 x 附近是 Lipschitz 连续的, 则对所有的 $h \in X$, 有 $f_-^{\downarrow}(x,h) = f'_-(x,h)$ 且 $f_+^{\downarrow}(x,h) = f'_+(x,h)$. 由 (2.68) 与 (2.69) 易证这一结论成立, 其证明类似于命题 2.49.

称 f 在 x 处沿方向 h 是方向上图可微的, 若 $f_-^{\downarrow}(x,h) = f_+^{\downarrow}(x,h)$. 此种情形记 $f^{\downarrow}(x,h)$ 为它们公共的值. 注意到, 即使 f 是凸函数, $f^{\downarrow}(x,h)$ 与 $f'(x,h)$ 也可能是不同的 (见例 2.67).

若 f 是方向可微的, 其二阶方向导数定义为

$$f''(x\,;h,w) := \lim_{t\downarrow 0} \frac{f\left(x+th+\frac{1}{2}t^2w\right) - f(x) - tf'(x,h)}{\frac{1}{2}t^2}, \tag{2.71}$$

若上述极限存在 (文献中有时称 $f''(x\,;h,w)$ 是抛物 (parabolic) 的二阶方向导数). 上二阶方向导数与下二阶方向导数可以以类似的方式来定义, 如

$$f''_+(x\,;h,w) := \limsup_{t\downarrow 0} \frac{f\left(x+th+\frac{1}{2}t^2w\right) - f(x) - tf'(x,h)}{\frac{1}{2}t^2}.$$

注意, 若 X 是 Banach 空间且 f 具有二阶 Taylor 展式

$$f(x+h) = f(x) + Df(x)h + \frac{1}{2}D^2f(x)(h,h) + o(\|h\|^2), \tag{2.72}$$

则

$$f''(x\,;h,w) = Df(x)w + D^2f(x)(h,h). \tag{2.73}$$

设 $f(x)$ 与方向上图导数$f_-^{\downarrow}(x,h)$, $f_+^{\downarrow}(x,h)$ 是有限的, 也用到下二阶上图导数与上二阶上图导数

$$f_-^{\downarrow\downarrow}(x\,;h,\cdot) := \mathrm{e} - \liminf_{t\downarrow 0} \frac{f\left(x+th+\frac{1}{2}t^2\right) - f(x) - tf_-^{\downarrow}(x,h)}{\frac{1}{2}t^2}, \tag{2.74}$$

$$f_{+}^{\downarrow\downarrow}(x\,;h,\cdot) := \mathrm{e}-\limsup_{t\downarrow 0}\frac{f\left(x+th+\dfrac{1}{2}t^2\right)-f(x)-tf_{+}^{\downarrow}(x,h)}{\dfrac{1}{2}t^2},\tag{2.75}$$

也可表示为

$$f_{-}^{\downarrow\downarrow}(x\,;h,w) = \liminf_{\substack{t\downarrow 0\\ w'\to w}}\frac{f\left(x+th+\dfrac{1}{2}t^2w'\right)-f(x)-tf_{-}^{\downarrow}(x,h)}{\dfrac{1}{2}t^2},\tag{2.76}$$

$$f_{+}^{\downarrow\downarrow}(x;h,w)=\sup_{\{t_n\}\in\varSigma}\left[\liminf_{\substack{n\to\infty\\ w'\to w}}\frac{f\left(x+t_nh+\dfrac{1}{2}t_n^2w'\right)-f(x)-t_nf_{+}^{\downarrow}(x,h)}{\dfrac{1}{2}t_n^2}\right].\tag{2.77}$$

称 f 在 x 处沿方向 h 是二阶方向上图可微的, 若 $f_{-}^{\downarrow\downarrow}(x\,;h,\cdot)=f_{+}^{\downarrow\downarrow}(x\,;h,\cdot)$. 再次注意到, 若 $f(\cdot)$ 是 Lipschitz 连续的, 在 x 处方向可微, 则对 $h,w\in X$ 有 $f_{-}^{\downarrow\downarrow}(x\,;h,w)=f''_{-}(x\,;h,w)$ 且 $f_{+}^{\downarrow\downarrow}(x\,;h,w)=f''_{+}(x\,;h,w)$.

类似于 (2.71) 的方式, 可以定义映射 $g:X\to Y$ 的二阶方向导数, 其中 X 与 Y 是 Banach 空间. 即称 g 在 $x\in X$ 处沿方向 $h\in X$ 二阶方向可微, 若 $g'(x,h)$ 和极限

$$g''(x\,;h,w) := \lim_{t\downarrow 0}\frac{g\left(x+th+\dfrac{1}{2}t^2w\right)-g(x)-tg'(x,h)}{\dfrac{1}{2}t^2}\tag{2.78}$$

对所有的 $w\in X$ 均存在. 进一步, 若

$$g''(x\,;h,w) := \lim_{\substack{t\downarrow 0\\ w'\to w}}\frac{g\left(x+th+\dfrac{1}{2}t^2w'\right)-g(x)-tg'(x,h)}{\dfrac{1}{2}t^2},\tag{2.79}$$

则称 g 在 $x\in X$ 处沿方向 $h\in X$ 是 Hadamard 意义下二阶方向可微的. 类似于命题 2.49, 不难证明, 若 g 在 x 的邻域内是 Lipschitz 连续的, 在 $x\in X$ 处沿方向 $h\in X$ 是二阶方向可微的, 则它是 Hadamard 意义下二阶方向可微的. 类似于命题 2.47, 有下述链式法则.

命题 2.53 设 X,Y,Z 是 Banach 空间, $g:X\to Y$, $f:Y\to Z$. 设 g 在 $x\in X$ 处沿方向 $h\in X$ 是二阶方向可微的, f 在 $g(x)$ 处是 Hadamard 方向可微的, 它在 $g(x)$ 沿方向 $g'(x,h)$ 是 Hadamard 二阶方向可微的. 则复合映射 $f\circ g:X\to Z$ 在 $x\in X$ 处沿方向 $h\in X$ 是二阶方向可微的, 且

$$(f\circ g)''(x\,;h,w) := f''(g(x)\,;g'(x,h),g''(x\,;h,w)).\tag{2.80}$$

证明　首先, 由命题 2.47 得, 对一阶方向导数相应的链式法则 (2.39) 成立. 进一步, 由 (2.78) 得

$$g\left(x+th+\frac{1}{2}t^2w\right)=g(x)+tg'(x,h)+\frac{1}{2}t^2g''(x\,;h,w)+o(t^2).$$

由 f 的 Hadamard 二阶方向可微性的定义可得 (2.80). □

尤其, 若 g 在 x 处是二阶连续可微的, 则它在 x 处是二阶方向可微的, 且

$$g''(x\,;h,w)=Dg(x)w+D^2g(x)(h,h). \tag{2.81}$$

此种情况下, 在上述命题的假设之下, 公式 (2.80) 具有下述形式

$$(f\circ g)''(x\,;h,w)=f''(g(x)\,;Dg(x)h,Dg(x)w+D^2g(x)(h,h)). \tag{2.82}$$

在 3.2.1 节的命题 3.42 中, 对增广实值的凸函数f, 用不同的方法可推导出类似的链式法则.

2.2.4　切锥

这一节, 设 X 与 Y 是局部凸的拓扑向量空间, S 是 X 的闭子集. 引入 S 在 $x\in S$ 处的切锥与法锥的概念. 若 X 是 Banach 空间, 用 $\operatorname{dist}(x,S)$ 记 $x\in X$ 到 S 距离, 即 $\operatorname{dist}(x,S):=\inf_{z\in S}\|x-z\|$. 这里用 Painlevé Kuratowski 意义下 (见 (2.56)~(2.59)) 的上集合极限与下集合极限的定义来描述.

定义 2.54　对 $S\subset X$, $x\in S$, 定义下述集合:

雷达锥(the radial cone)

$$\mathcal{R}_S(x):=\{h\in X:\exists\, t^*>0,\forall t\in[0,t^*],x+th\in S\}. \tag{2.83}$$

余切 (Bouligand) 锥 (the contingent(Bouligand) cone)

$$T_S(x):=\limsup_{t\downarrow 0}\frac{S-x}{t}. \tag{2.84}$$

内切锥(the inner tangent cone)

$$T_S^i(x):=\liminf_{t\downarrow 0}\frac{S-x}{t}. \tag{2.85}$$

Clarke 切锥(the Clarke tangent cone)

$$T_S^c(x):=\liminf_{\substack{t\downarrow 0\\ S\ni x'\to x}}\frac{S-x'}{t}. \tag{2.86}$$

用 $S \ni x' \to x$ 表示在 S 中 $x' \to x$. 不难看出, 由 (2.83)~(2.86) 定义的集合实际上均是锥. 后续的讨论主要处理余切和内切锥. 注意, 尽管这些切锥的重要性是众所周知的, 但在文献中, 上述的术语并不是统一的. 若 x 不属于 S, 约定这些锥为空集. 显然有

$$\mathcal{R}_S(x) \subset T_S^i(x) \subset T_S(x).$$

余切锥、内切锥与 Clarke 切锥是闭的, 而可能发生雷达锥非闭的情形. 若 X 是 Banach 空间, 余切锥与内切锥可以写为下述形式

$$T_S(x) = \{h \in X : \exists t_n \downarrow 0, \operatorname{dist}(x + t_n h, S) = o(t_n)\}, \tag{2.87}$$

$$T_S^i(x) = \{h \in X : \operatorname{dist}(x + th, S) = o(t), t \geqslant 0\}. \tag{2.88}$$

显然, Clarke 切锥$T_S^c(x)$ 是由这样的 $h \in X$ 组成: 即对任何 $t_n \downarrow 0$, 对任何 $x_n \to x$, $x_n \in S \setminus \{x\}$, 存在 $y_n \in S$, 满足 $(y_n - x_n)/t_n \to h$. 等价地, $h \in T_S^c(x)$ 当且仅当对任何序列 $t_n \downarrow 0$, $x_n \to x$, $x_n \in S \setminus \{x\}$, 存在收敛到 h 的序列 h_n 满足 $x_n + t_n h_n \in S$.

一般地, 这些锥是不同的, 雷达锥、余切锥与内切锥可以是非凸的. 然而, 对凸集而言, 余切锥、内切锥与 Clarke 切锥相等, 都等于雷达锥的闭包, 是凸的.

命题 2.55 若 S 是凸闭集, $x \in S$, 则

$$\mathcal{R}_S(x) := \bigcup_{t>0} \{t^{-1}(S - x)\} \tag{2.89}$$

且

$$T_S(x) = T_S^i(x) = \operatorname{cl}[\mathcal{R}_S(x)]. \tag{2.90}$$

证明 由 S 的凸性, 公式 (2.89) 是显然的结果. S 的凸性可推得, $t^{-1}(S - x)$ 是 t 的递减函数 (在集合包含意义下). 因此, 当 $t \downarrow 0$ 时, $t^{-1}(S - x)$ 的上极限与下极限相等, 是 $\mathcal{R}_S(x)$ 的拓扑包. □

由上述命题得, 对凸集而言, 余切锥与内切锥相等, 且等价地 (在 X 是 Banach 空间的情形)

$$T_S(x) = \{h \in X : \operatorname{dist}(x + th, S) = o(t), t \geqslant 0\}. \tag{2.91}$$

因此, 对凸集而言, 我们称其为切锥, 而不特指余切或内切锥.

注意到, 若 $x \in S$, 则 $0 \in T_S(x)$ 且 $0 \in T_S^i(x)$, 因此这两个锥是非空的. 另一方面, 若 x 是 S 的孤立点 (即存在 x 的邻域 N 满足 $S \cap N = \{x\}$), 则不存在收敛到 x 的序列 $x_n \in S \setminus \{x\}$, 因此 Clarke 切锥 $T_S^c(x)$ 是空集. 当然, 若集合 S 是凸集且非单点集, 则它没有孤立点. 此种情况, 可证明 Clarke 切锥也等于余切锥和内切锥(见命题 2.57).

命题 2.56 设 S 是 Banach 空间 X 的闭子集 (不一定是凸的), $x \in S$, 则 Clarke 切锥 $T_S^c(x)$ 是凸的, $T_S^c(x) \subset T_S^i(x)$, 且

$$T_S^c(x) \subset \liminf_{S \ni x' \to x} T_S(x'). \tag{2.92}$$

进一步, 若空间 X 是有限维的, 则

$$T_S^c(x) = \liminf_{S \ni x' \to x} T_S(x'). \tag{2.93}$$

证明 若 x 是 S 的孤立点, 则 $T_S^c(x)$ 是空集, 上述结论是平凡成立的. 因此, 设 x 不是 S 的孤立点. 由存在收敛到 x 的序列 $\bar{x}_n \in S \setminus \{x\}$, 有 $h \in T_S^c(x)$ 当且仅当对任何 $t_n \downarrow 0$, $x_n \to x$, $x_n \in S \setminus \{x\}$, 存在序列 $y_n \in S$, 满足 $(y_n - x_n)/t_n \to h$. 因此设 $t_n \downarrow 0$ 是给定序列, 令 $\{x_n'\}$ 是序列 $\{\bar{x}_n\}$ 的子序列, 满足 $(x_n' - x)/t_n \to 0$. 若 $h \in T_S^c(x)$, 则存在 $y_n \in S$, 满足 $(y_n - x_n')/t_n \to h$. 从而有 $(y_n - x)/t_n \to h$, 因此 $h \in T_S^i(x)$(因序列 $t_n \downarrow 0$ 是任意的). 这证得包含关系 $T_S^c(x) \subset T_S^i(x)$.

因为 $T_S^c(x)$ 是锥, 为证 $T_S^c(x)$ 是凸的, 只需证, 若 h_1, $h_2 \in T_S^c(x)$, 则 $h_1 + h_2 \in T_S^c(x)$. 考虑序列 $t_n \downarrow 0$, $x_n \to x$, $x_n \in S \setminus \{x\}$. 由于 $h_1 \in T_S^c(x)$, 存在序列 $y_n \in S \setminus \{x\}$, 满足 $(y_n - x_n)/t_n \to h_1$. 又由于 $h_2 \in T_S^c(x)$, 则存在序列 $z_n \in S$, 满足 $(z_n - y_n)/t_n \to h_1$①. 于是有 $(z_n - x_n)/t_n \to h_1 + h_2$, 因此有 $h_1 + h_2 \in T_S^c(x)$. 这表明 $T_S^c(x)$ 是凸的.

现在证 (2.92). 用 $\mathcal{T}(x)$ 记 (2.92) 右端的集合. 需证明若 $h \notin \mathcal{T}(x)$, 则 $h \notin T_S^c(x)$. 由 $\mathcal{T}(x)$ 与 $T_S^c(x)$ 的定义. $h \notin \mathcal{T}(x)$ 当且仅当

$$\exists \varepsilon > 0, \exists x_n \to x, x_n \in S \setminus \{x\} : \operatorname{dist}(h, T_S(x_n)) \geqslant \varepsilon, \tag{2.94}$$

而 $h \notin T_S^c(x)$ 当且仅当

$$\exists \varepsilon > 0, \exists x_n \to x, x_n \in S \setminus \{x\}, \exists t_n \downarrow 0 : (x_n + t_n \overline{B}(h, \varepsilon)) \cap S = \varnothing. \tag{2.95}$$

设 $h \notin \mathcal{T}(x)$. 由 (2.94), 存在 $\varepsilon > 0$, $S \ni x_n \to x$ 满足 $B(h, \varepsilon) \cap T_S(x_n) = \varnothing$. 对 $\varepsilon' \in (0, \varepsilon)$, 存在 $t_n' \downarrow 0$, 对 $t \in (0, t_n')$, 集合 $B(h, \varepsilon') \cap [t^{-1}(S - x_n)]$ 是空集, 这意味着 $(x_n + tB(h, \varepsilon')) \cap S = \varnothing$. 取 $t_n \in (0, t_n')$, 有 $t_n \downarrow 0$, 由 (2.95) 得 $h \notin T_S^c(x)$. 这就证得包含关系 (2.92).

相反地, 设 X 是有限维空间, $h \notin T_S^c(x)$, 设 ε, x_n 与 t_n 满足 (2.95). 为验证 (2.94), 只需证明, 若对 $x' \in S$, $\tau > 0$, 集合 $(x' + \tau \overline{B}(h, \varepsilon)) \cap S$ 是空集, 则存在 $\bar{x} \in S$, 满足 $\operatorname{dist}(h, T_S(\bar{x})) \geqslant \varepsilon/2$ 且 $\|x' - \bar{x}\| \leqslant \tau(\|h\| + \varepsilon)$. 因为 X 是有限维的, 不难由紧

① 原文为 $(z_n - y_n)/t_n \to h_1$.

致性结论得, 满足 $(x'+t\overline{B}(h,\varepsilon))\cap S\neq\varnothing$ 的 $t\in[0,\tau]$ 的集合是闭的, 由假设知, 这一集合不包含 τ. 因为 $x'\in S$, 这一集合包含 0, 因而这一集合非空. 令 $\bar{t}$ 是此集合的最大元素. 有 $0\leqslant\bar{t}<\tau$, 因而 $t^*:=\tau-\bar{t}$ 是正的. 令 $\bar{x}$ 是集合 $(x'+\bar{t}\,\overline{B}(h,\varepsilon))\cap S$ 的元素. 可得到

$$\|x'-\bar{x}\|\leqslant\bar{t}(\|h\|+\varepsilon)\leqslant\tau(\|h\|+\varepsilon).$$

剩下还需证明对任何 $t\in(0,t^*)$, 集合 $B(h,\varepsilon)\cap[t^{-1}(S-\bar{x})]$ 是空集, 这意味着 $(\bar{x}+tB(h,\varepsilon))\cap S=\varnothing$. 由于 $\bar{x}\in x'+\bar{t}\,\overline{B}(h,\varepsilon)$, 对任何 $t\in(0,t^*)$, 有

$$\bar{x}+tB(h,\varepsilon)\subset x'+(\bar{t}+t)\overline{B}(h,\varepsilon)=x'+t'\overline{B}(h,\varepsilon),$$

其中 $t':=\bar{t}+t\in(\bar{t},\tau)$. 由构造得到, 对任何 $t'\in(\bar{t},\tau)$, 集合 $x'+t'\overline{B}(h,\varepsilon)$ 与集合 S 没有公共点. 这表明, 对任何 $t\in(0,t^*)$, 实际上均有 $(\bar{x}+tB(h,\varepsilon))\cap S$ 是空集. 这就完成了证明. □

命题 2.57 设 S 是 Banach 空间X 的非空凸闭子集, $x\in S$, 设 S 不是单点集. 则 $T_S^{\mathrm{c}}(x)=T_S(x)$.

证明 由命题 2.56 得 $T_S^{\mathrm{c}}(x)\subset T_S(x)$. 因此, 只需证明包含关系 $T_S(x)\subset T_S^{\mathrm{c}}(x)$. 因为 $T_S(x)$ 是 $\mathcal{R}_S(x)$ 的拓扑包且 $T_S^{\mathrm{c}}(x)$ 是闭集, 只需要证明 $\mathcal{R}_S(x)\subset T_S^{\mathrm{c}}(x)$. 考虑 $h\in\mathcal{R}_S(x)$. 则存在 $\tau>0$, 满足 $x+\tau h\in S$. 考虑某序列 $x_n\to x, x_n\in S\setminus\{x\}, t_n\downarrow 0$. 则对 $h_n:=\tau^{-1}(x+\tau h-x_n)$, 有 $h_n\to h$ 且

$$x_n+t_nh_n=t_n\tau^{-1}(x+\tau h)+(1-t_n\tau^{-1})x_n,$$

因而由 S 的凸性, 由 $t_n\leqslant\tau$, $x_n+t_nh_n\in S$. 这可推出 $h\in T_S^{\mathrm{c}}(x)$, 证毕. □

由上述命题可知, 若 S 是 Banach 空间X 的凸的闭子集, 且 S 不是单点集, 则 S 的切锥具有下述半连续性的重要性质

$$T_S(x)\subset\liminf_{S\ni x'\to x}T_S(x'). \tag{2.96}$$

若 X 还是有限维的, 则 $T_S(x)$ 等于 (2.96) 的右端极限.

余切锥$T_S(x)$ 的极锥被称为 S 在 x 处的法锥(normal cone). 即

$$N_S(x):=[T_S(x)]^-. \tag{2.97}$$

若 S 是凸集, 因为 $T_S(x)$ 是 $\mathcal{R}_S(x)$ 的闭包, 得到 $N_S(x):=[\mathcal{R}_S(x)]^-$. 因此, 若 $x\in S$ 且 S 是凸集, 则

$$N_S(x)=\{x^*\in X^*:\langle x^*,z-x\rangle\leqslant 0,\ \forall z\in S\}. \tag{2.98}$$

由定义, 若 $x\notin S$, $N_S(x)=\varnothing$.

注意到, 若 X^* 是 Banach 空间X 的对偶, K^* 是 X^* 的凸子集, 则在弱 * 拓扑下 $\mathcal{R}_{K^*}(x^*)$ 的拓扑包可不同于 X^* 的强拓扑下的$\mathcal{R}_{K^*}(x^*)$ 的拓扑包, 见例 2.38.

由 (2.98) 可得, 法锥多值函数在下述意义下是单调的(monotone), 即对任何 x_1, $x_2 \in X$, 对任何 $x_1^* \in N_S(x_1)$, $x_2^* \in N_S(x_2)$,

$$\langle x_1^* - x_2^*, x_1 - x_2 \rangle \geqslant 0. \tag{2.99}$$

考虑增广实值 (不必凸) 函数 $f : X \to \overline{\mathbb{R}}$ 与集合 $K := \operatorname{epi} f$. 回顾 f 是下半连续的当且仅当集合 $\operatorname{epi}(f)$ 是 $X \times \mathbb{R}$ 的闭子集.

命题 2.58　设 $f : X \to \overline{\mathbb{R}}$ 是增广实值函数, 令 $x \in X$, 使 $f(x)$ 是有限的. 则

$$T_{\operatorname{epi} f}(x, f(x)) = \operatorname{epi} f_-^{\downarrow}(x, \cdot), \tag{2.100}$$

$$T_{\operatorname{epi} f}^i(x, f(x)) = \operatorname{epi} f_+^{\downarrow}(x, \cdot). \tag{2.101}$$

证明　由定义得, $f_-^{\downarrow}(x, \cdot)$ 的上图与 $t \downarrow 0$ 时, $\operatorname{epi}\{(f(x + t\cdot) - f(x))/t\}$ 的上集合极限重合. 连同余切锥的定义 (2.84), 可推出 (2.100). 等式 (2.101) 的证明是类似的. □

注 2.59　上述命题的结果可以通过下述可交换的图示表示 (以上上图导数为例)

$$\begin{array}{ccc} f & \xrightarrow{\text{epi}} & \operatorname{epi} f \\ \downarrow & & \downarrow \\ f_+^{\downarrow} & \xrightarrow{\text{epi}} & \operatorname{epi} f_+^{\downarrow} \end{array}$$

其中竖直箭头表示计算上上图导数[①](在左面) 与内切锥(在右面) 运算.

增广实值函数是凸的充分必要条件是它的上图是凸的. 因为对凸集而言, 余切与内切锥是相同的, 且是凸的, 由上述命题可得, 若 f 是凸的, 则方向上图导数$f^{\downarrow}(x, \cdot) = f_-^{\downarrow}(x, \cdot) = f_+^{\downarrow}(x, \cdot)$ 存在且是凸的.

令 $f : X \to \overline{\mathbb{R}}$ 是凸函数, 在点 $x \in X$ 处取有限值. 由 f 的凸性得, 对任意的 $h \in X$, 差商 $[f(x + th) - f(x)]/t$ 当 t 递减到 0 是不递增的. 我们得到

$$\lim_{t \downarrow 0} \frac{f(x + th) - f(x)}{t} = \inf_{t > 0} \frac{f(x + th) - f(x)}{t}, \tag{2.102}$$

其中 (2.102) 左端的极限总是存在的, 尽管它可能取 $-\infty$ 或 $+\infty$. 因此, 称 f 在点 x 处是方向可微的, 且

$$f'(x, h) = \inf_{t > 0} \frac{f(x + th) - f(x)}{t}. \tag{2.103}$$

① 原著作中为下上图导数.

方向导数$f'(x,\cdot)$ 是正齐次的, 即对任何 $h \in X$, $t > 0$, 有 $f'(x,th) = tf'(x,h)$. 因为 f 是凸函数, 方向导数也是凸的.

命题 2.60 设 $f : X \to \overline{\mathbb{R}}$ 是凸函数, 令 $x \in X$ 是一点, 满足 $f(x)$ 有限. 则 $f'(x,\cdot)$ 的上图的闭包与 epi$f$ 在 $(x, f(x))$ 处的切锥重合, 即

$$\mathrm{cl}[\mathrm{epi}f'(x,\cdot)] = T_{\mathrm{epi}f}(x, f(x)). \tag{2.104}$$

证明 不失一般性, 不妨设 $x = 0$, $f(x) = 0$. 设 $(h,c) \in \mathrm{epi}f'(0,\cdot)$, 即 $c \geqslant f'(0,h)$. 则对任何 $c' > c$, 对充分小的 t, 有 $t(h,c) \in \mathrm{epi}f$, 即 $(h,c') \in \mathcal{R}_{\mathrm{epi}f}(x, f(x))$. 由切锥的闭性, 得到 $(h,c) \in T_{\mathrm{epi}f}(x, f(x))$, 因此有

$$\mathrm{cl}[\mathrm{epi}f'(x,\cdot)] \subset T_{\mathrm{epi}f}(x, f(x)). \tag{2.105}$$

相反地, 设 $(h,c) \in \mathcal{R}_{\mathrm{epi}f}(x, f(x))$. 这就意味着对充分小的 $t > 0$, $tc \geqslant f(th)$. 结果有 $c \geqslant f'(x,h)$, 即有 $(h,c) \in \mathrm{epi}f'(x,\cdot)$. 则与 (2.105) 相反的包含关系成立. 结论得证. □

由命题 2.58 与命题 2.60 得, 对凸函数f, $f'(x,\cdot)$ 的下半连续包即是 $f^{\downarrow}(x,\cdot)$. 这意味着

$$\liminf_{\substack{t\downarrow 0\\ h'\to h}} \frac{f(x+th') - f(x)}{t} = \min\left\{\liminf_{h'\to h} f'(x,h'), f'(x,h)\right\}. \tag{2.106}$$

下述命题给出了切集非常有用的刻画. 回顾 $g^{\downarrow}(x_0,h)$ 记增广实值函数 $g(x)$ 的方向上图函数, 凸函数总是方向上图可微的.

命题 2.61 设 $g : X \to \overline{\mathbb{R}}$ 是下半连续凸函数. 考虑相联系的水平集 $S := \{x \in X : g(x) \leqslant 0\}$. 设 $g(x_0) = 0$, 且存在 $\bar{x}$ 满足 $g(\bar{x}) < 0$(Slater 条件), 则

$$T_S(x_0) = \left\{h \in X : g^{\downarrow}(x_0,h) \leqslant 0\right\}. \tag{2.107}$$

证明 由于 g 是凸函数, 集合 S 是凸的, 又因为 g 是下半连续的, 集合 S 是闭集. 注意到下述两个包含关系成立

$$\left\{h \in X : g^{\downarrow}(x_0,h) < 0\right\} \subset T_S(x_0), \tag{2.108}$$

$$T_S(x_0) \subset \left\{h \in X : g^{\downarrow}(x_0,h) \leqslant 0\right\}. \tag{2.109}$$

事实上, 若 $g^{\downarrow}(x_0,h) < 0$, 则由 (2.68), 存在 $t_n \downarrow 0$ 及 $h_n \to h$, 满足 $g(x_0 + t_n h_n) < 0$. 结果 $h_n \in \mathcal{R}_S(x_0)$, 从而 $h \in T_S(x_0)$. 这证得 (2.108). 令 $h \in T_S(x_0)$. 则存在序列 $\{h_n\} \subset \mathcal{R}_S(x_0)$, 满足 $h_n \to h$. 因为 $h_n \in \mathcal{R}_S(x_0)$, 存在 $t_n > 0$, $x_0 + t_n h_n \in S$, 因而有 $g(x_0 + t_n h_n) \leqslant 0$. 取极限得到 $g^{\downarrow}(x_0,h) \leqslant 0$, 这就证得 (2.109).

只需证明, (2.108) 左端的拓扑包与 (2.109) 右端相同. 为此需要用到 Slater 条件. 考虑向量 $h \in X$ 满足 $g^{\downarrow}(x_0, h) \leqslant 0$. 则存在序列 $h_n \to h$, $\varepsilon_n \downarrow 0$ 及 $t_n \downarrow 0$, 满足 $g(x_0 + t_n h_n) \leqslant \varepsilon_n t_n$. 由 g 的凸性得到, 对所有的 $t \in [0, t_n]$, 有 $g(x_0 + th_n) \leqslant \varepsilon_n t$. 对满足 $\varepsilon_n/\alpha_n \to 0$ 的序列 $\alpha_n \downarrow 0$, 置

$$h_n' := \alpha_n(\bar{x} - x_0) + (1 - \alpha_n)h_n, \quad \beta_{n,t} := (1 - t\alpha_n)^{-1}(1 - \alpha_n).$$

由于

$$x_0 + th_n' = t\alpha_n\bar{x} + (1 - t\alpha_n)x_0 + t(1 - \alpha_n)h_n = (1 - t\alpha_n)(x_0 + t\beta_n h_n) + t\alpha_n\bar{x},$$

由 g 的凸性, 对 $t > 0$ 充分小, 有

$$g(x_0 + th_n') \leqslant (1 - t\alpha_n)g(x_0 + t\beta_n h_n) + t\alpha_n g(\bar{x}).$$

对固定的 n, 对充分小的 $t > 0$, 有 $g(x_0 + t\beta_n h_n) \leqslant t\beta_n\varepsilon_n$. 得到 $g^{\downarrow}(x_0, h_n') \leqslant (1 - \alpha_n)\varepsilon_n + \alpha_n g(\bar{x})$. 因为 $\varepsilon_n/\alpha_n \to 0$, 这意味着 $g^{\downarrow}(x_0, h_n') < 0$, 因此 h_n' 属于 (2.108) 的左端. 因为 $h_n' \to h$, 得到结论. □

如果上述命题中函数 g 是凸函数, 在 x_0 处是连续的, 则它在 x_0 的附近是 Lipschitz 连续的且在 x_0 处是方向可微的 (见 2.4 节). 此种情况下, $g^{\downarrow}(x_0, \cdot) = g'(x_0, \cdot)$, 因此, 在 Slater 条件下, 切锥 $T_S(x_0)$ 与集合 $\{h : g'(x_0, h) \leqslant 0\}$ 是重合的. 注意到, 命题 2.61 中的 Slater 条件是本质性的. 如考虑 $g(x) = x^2$, $x \in \mathbb{R}$, $x_0 = 0$. 则 $K := \{x : g(x) \leqslant 0\}=\{0\}$, 因此 $T_K(0) = \{0\}$. 另一方面, $Dg(0) = 0$, 则 $\{h : g'(0, h) \leqslant 0\} = \mathbb{R}$.

例 2.62　令 K 是 X 的闭凸锥, $x \in K$. 则 $\mathcal{R}_K(x) = K + [\![x]\!]$, 其中 $[\![x]\!] := \{tx : t \in \mathbb{R}\}$ 记由向量 x 生成的线性空间. 实际上, 显然有 $K \subset \mathcal{R}_K(x)$ 且 $[\![x]\!] \subset \mathcal{R}_K(x)$. 由于 $\mathcal{R}_K(x)$ 是凸锥, 有 $K + [\![x]\!] \subset \mathcal{R}_K(x)$. 相反地, 由于 $\mathcal{R}_K(x) = \mathbb{R}_+(K - x) \subset K + [\![x]\!]$, 有 $\mathcal{R}_K(x) \subset K + [\![x]\!]$.

由 (2.31), 由 $\mathcal{R}_K(x) = K + [\![x]\!]$ 得

$$N_K(x) = K^- \cap [\![x]\!]^{\perp} = \{x^* \in K^- : \langle x^*, x\rangle = 0\}. \tag{2.110}$$

例 2.63　令 Ω 是紧致的 Harsdorff 拓扑空间, 考虑 Banach 空间$X := C(\Omega)$(见例 2.37). $C(\Omega)$ 的对偶空间是 Ω 上的 (有限符号的) 正则 Borel 测度构成的线性空间. 考虑非正值的连续函数构成的锥 $K := C_-(\Omega)$,

$$C_-(\Omega) := \{x \in C(\Omega) : x(\omega) \leqslant 0, \forall\omega \in \Omega\}. \tag{2.111}$$

锥 K 的极锥由 Ω 上的非负正则 Borel 测度集构成, 即 $\mu \in K^-$ 当且仅当对任何 $A \in \mathcal{B}$ 有 $\mu(A) \geqslant 0$. 事实上, 令 $\mu \in K^-$, 考虑 Jordan 分解 $\mu = \mu^+ - \mu^-$, 其中 μ^+

与 μ^- 分别是 μ 的正的与负的部分. 回顾, μ^+ 与 μ^- 是集中在不相交的 Borel 集 Σ_1 与 Σ_2 上的非负的正则的 Borel 测度. 再考虑全差分测度 $|\mu| := \mu^+ + \mu^-$. 用反证法证明. 假设测度 μ 的负部 μ^- 非零, 即 $\mu^-(\Sigma_2) = \alpha > 0$. 则由于 μ 是正则的, 可找到闭集 $A \subset \Omega$ 与开集$B \subset \Omega$ 满足 $A \subset \Sigma_2 \subset B$ 且 $|\mu|(B \setminus A) < \alpha/2$. 进一步, 存在连续函数 $y \in C(\Omega)$, 满足对所有的 $\omega \in A$, 有 $y(\omega) = -1$, 对所有的 $\omega \in \Omega \setminus B$, $y(\omega) = 0$, 对所有的 $\omega \in \Omega$, 有 $-1 \leqslant y(\omega) \leqslant 0$. 得到 $\mu(A) < -\alpha/2$ 且 $\mu(B \setminus A) < \alpha/2$, 因此

$$\langle \mu, y \rangle = \int_\Omega y(\omega) \mathrm{d}\mu(\omega) > 0.$$

因为 $y \in C_-(\Omega)$, 这与 $\mu \in K^-$ 矛盾.

对函数 $x \in K$, 记

$$\Delta(x) := \{\omega \in \Omega : x(\omega) = 0\} \tag{2.112}$$

为 x 的接触点的集合. 对 $x \in K$, 有

$$T_K(x) = \{y \in C(\Omega) : y(\omega) \leqslant 0, \forall \omega \in \Delta(x)\}. \tag{2.113}$$

事实上, 因为 K 是锥, 由例 2.62 得 $T_K(x) = \mathrm{cl}(K + [\![x]\!])$. 显然, 若 $y \in K + [\![x]\!]$, 则对 $\omega \in \Delta(x)$, 有 $y(\omega) \leqslant 0$. 由连续性, 对 $y \in T_K(x)$ 也有同样的结论. 这表明 $T_K(x)$ 包含在 (2.113) 的右端. 相反地, 令 $y \in C(\Omega)$ 满足对 $\omega \in \Delta(x)$, 有 $y(\omega) < 0$. 则由于 Ω 是紧致的, $\Delta(x)$ 是紧致的, 对充分小的 $t > 0$, 有 $x + ty \in C_-(\Omega)$, 因而 $y \in \mathcal{R}_K(x)$. 因为切锥 $T_K(x)$ 是闭的(以 $C(\Omega)$ 的范数拓扑), 所以 (2.113) 中的右端包含在 $T_K(x)$, 这证得结论.

不难由 (2.113) 推得

$$N_K(x) = \{\mu \in C(\Omega)^* : \mu \geqslant 0, \mathrm{supp}(\mu) \subset \Delta(x)\}. \tag{2.114}$$

例 2.64 令 $X := L_p(\Omega, \mathcal{F}, \mu)$, 或简言之, $X = L_p(\Omega)$, 其中 $(\Omega, \mathcal{F}, \mu)$ 是测度空间, $p \in [1, \infty)$(例 2.39). 设 $K \subset L_p(\Omega)$ 是由 a.e.(几乎处处) 非正值的函数组成, 记为 $K = [L_p(\Omega)]_-$. 回顾 p 的共轭数q 满足 $1/p + 1/q = 1$. 则 K 是闭凸锥, 它的极锥K^{-1} 是由 $L_q(\Omega)$ 的 a.e. 非负值函数构成. 锥 K 与 K^- 均有空的内部.

不难证明 (留为练习), K 在 $x \in K$ 处的切锥是

$$T_K(x) = \Big\{h \in X : h(\omega) \leqslant 0, \text{ 若几乎处处 } x(\omega) = 0\Big\}. \tag{2.115}$$

尤其, 若对几乎处处的 ω, $x(\omega) < 0$, 则 $T_K(x) = X$. 这可与 (2.113) 给出的相应的切锥相比较, 其中 $x(\omega)$ 视为 $C(\Omega)$ 的元素.

当 $p = 2$ 时, $X := L_p(\Omega, \mathcal{F}, \mu)$ 成为 Hilbert 空间, 其数乘积为

$$\langle y, x \rangle = \int_\Omega y(t) x(t) \mathrm{d}t, \tag{2.116}$$

且 X^* 等同于 X. 此种情况下, $K^- = -K$.

例 2.65　下述例子与半定规划相关. 令 $\mathcal{S}^p$ 是 $p \times p$ 对称矩阵构成的线性空间, 赋予数乘积

$$A \cdot B := \text{trace}AB,$$

令 $K := \mathcal{S}_-^p$ 为 $p \times p$ 的负半定对称矩阵构成的锥. 它的极锥是由 $p \times p$ 的正半定对称矩阵构成的锥 $\mathcal{S}_+^p$. 注意到锥 K 可以表示为下述形式

$$\mathcal{S}_-^p = \{A \in \mathcal{S}^p : \lambda_{\max}(A) \leqslant 0\}, \tag{2.117}$$

其中 $\lambda_{\max}(A)$ 即矩阵 $A \in \mathcal{S}^p$ 的最大特征值.

这一最大特征值可表示为

$$\lambda_{\max}(A) = \max_{\|x\| \leqslant 1} x^{\mathrm{T}} A x.$$

结果, 由于 $\lambda_{\max}(A)$ 可写为关于 A 线性函数的最大值函数, 它是凸函数. 为计算方向导数, 可用 $\lambda_{\max}(A)$ 的上述最大值函数的表达式 (见例 4.20). 令 $E = [e_1, \cdots, e_s]$ 是 $p \times s$ 矩阵, 其列 $e_1, \cdots, e_s$ 由对应最大特征值 $\lambda_{\max}(A)$ 的特征向量空间的一组基构成. 则 $\lambda'_{\max}(A, H)$ 等于 $s \times s$ 矩阵 $E^{\mathrm{T}} H E$ 的最大特征值.

不难看到, Slater 条件对函数 $\lambda_{\max}(\cdot)$ 是成立的. 设 $\lambda_{\max}(A) = 0$, 则 $\text{rank}(A) = p - s$, 其中 s 是 A 的最大特征值的重数. 则由命题 2.61 得

$$T_{\mathcal{S}_-^p}(A) = \{H \in \mathcal{S}^p : \lambda'_{\max}(A, H) \leqslant 0\}, \tag{2.118}$$

结合上面关于 $\lambda'_{\max}(A, H)$ 的公式, 有

$$T_{\mathcal{S}_-^p}(A) = \{H \in \mathcal{S}^p : E^{\mathrm{T}} H E \preceq 0\}, \tag{2.119}$$

注意到, 若 $\lambda_{\max}(A) < 0$, 则 A 是锥 $\mathcal{S}_-^p$ 的内部点, 此种情形 $T_{\mathcal{S}_-^p}(A) = \mathcal{S}^p$.

例 2.66　设 X_1, X_2 是 Banach 空间, 令 $S_1 \subset X_1$, $S_2 \subset X_2$ 是闭凸集. 考虑空间 $X = X_1 \times X_2$, 赋予积范数 $\|(x_1, x_2)\| := \|x_1\| + \|x_2\|$, 集合 $S := S_1 \times S_2 \subset X$. 得

$$\text{dist}\,((x_1, x_2), S_1 \times S_2) = \text{dist}(x_1, S_1) + \text{dist}(x_2, S_2), \tag{2.120}$$

从而

$$T_S((x_1, x_2)) = T_{S_1}(x_1) \times T_{S_2}(x_2). \tag{2.121}$$

例 2.67　令 A 是 X 的非空闭凸子集, $f(x) = I_A(x)$, A 的指示函数定义为

$$I_A(x) := \begin{cases} 0, & 若\ x \in A, \\ +\infty, & 若\ x \notin A. \end{cases} \tag{2.122}$$

注意到 f 是正常的下半连续凸函数. 考虑集合 $K := \text{epi}f = A \times \mathbb{R}_+$ 及点 $x \in A$. 则 $f'(x,h) = I_{\mathcal{R}_A(x)}(h)$, $f^{\downarrow}(x,h) = I_{T_A(x)}(h)$, 因此有 $\text{lsc}f'(x,\cdot) = f^{\downarrow}(x,\cdot)$. 进一步 (见例 2.66),

$$T_{\text{epi}f}(x,0) = T_A(x) \times \mathbb{R}_+ = \text{epi}I_{T_A(x)}(\cdot),$$

这些与命题 2.58 与命题 2.60 是一致的.

例 2.68 考虑 $f(x)$ 是有限个函数的最大值的实值函数的情形

$$f(x) = \max_{1 \leqslant i \leqslant n} f_i(x). \tag{2.123}$$

不难证明, 若每一个函数 f_i, $i = 1, \cdots, n$ 在 x 处是方向可微的, 则最大值函数 f 在 x 处亦是方向可微的, 且

$$f'(x,h) = \max_{i \in I(x)} f_i'(x,h), \tag{2.124}$$

其中

$$I(x) := \{i : \ f_i(x) = f(x), \ i \leqslant i \leqslant n\}.$$

若每一函数 f_i, $i = 1, \cdots, n$ 还均有二阶方向导数, 则最大值函数 f 也是二阶方向可微的, 且

$$f''(x\,;h,w) = \max_{i \in I_1(x,h)} f_i''(x;h,w), \tag{2.125}$$

其中

$$I_1(x,h) := \{i \in I(x) : \ f_i'(x,h) = \max_{i \in I(x)} f_i'(x,h)\}.$$

尤其, 若每一函数 f_i 是二阶可微的, 则最大值函数的二阶方向导数可结合 (2.73) 与 (2.125) 来计算. 后面还可以看到上述公式是如何推广到紧致集上最大值函数的情形的.

我们已经看到, 任何凸函数在取有限值点处是一阶方向可微的. 下面的例子表明, 并不是每一凸函数均是二阶方向可微的. 由定义易知, 若 f 是凸函数, 则上二阶方向导数$f_+''(x\,;h,\cdot)$ 也是凸的. 凸函数的下二阶方向导数 $f_-''(x\,;h,\cdot)$ 可能是非凸的 (见例 3.35).

例 2.69 先构造凸的分片线性函数 $y = \eta(x)$, $x \in \mathbb{R}$, 在两个抛物线 $y = x^2$ 与 $y = 2x^2$ 间振荡. 即以这样的方式构造 $\eta(x)$, 使它满足 $\eta(x) = \eta(-x)$, $\eta(0) = 0$, 且对某一单调递减收敛到 0 的序列 x_k, 函数在每一区间 $[x_{k+1}, x_k]$ 上是线性的, $\eta(x_k) = x_k^2$, 经过点 $(x_k, \eta(x_k))$ 与 $(x_{k+1}, \eta(x_{k+1}))$ 的直线与曲线 $y = 2x^2$ 相切. 这样的函数如何构造是非常显然的. 对给定的 $x_k > 0$, 考虑经过点 (x_k, x_k^2) 与曲线 $y = 2x^2$ 相切的直线. 直线与 $y = x^2$ 在 x_{k+1} 处相交. 可计算得 $x_{k+1} = (3 - 2\sqrt{2})x_k$. 显然,

$x_k > x_{k+1} > 0$, 因而可用迭代方式完成构造. 再注意到在点 $a_k := (1 - 1/\sqrt{2})x_k$, $\eta(a_k) = 2a_k^2$ 且 $a_{k+1} = (3 - 2\sqrt{2})a_k$.

现在定义 $f(x_1, x_2) := \eta(x_1) - x_2$. 函数 $f := \mathbb{R}^2 \to \mathbb{R}$ 是两个凸函数 $\eta(x_1)$ 与 $-x_2$ 的和, 因而亦是凸函数. 令 $h := (1, 0)$, $w := (0, 2)$. 则 $f(0) = 0$, $f'(0, h) = 0$ 且 $f\left(th + \frac{1}{2}t^2 w\right) = f(t, t^2) = \eta(t) - t^2$. 结果 $f\left(th + \frac{1}{2}t^2 w\right)/t^2$ 在 0 与 1 间振荡, 因而当 $t \downarrow 0$ 时没有极限. 因此得到 $f''(0; h, w)$ 是不存在的.

2.3 多值函数理论的若干结果

这一节讨论多值函数 (点到集映射) 相关的某些基本概念和结果. 令 X 与 Y 是 Banach 空间, 考虑多值函数 $\Psi : X \to 2^Y$, 映 X 到 Y 的子集集合. Ψ 的定义域, 值域分别定义为

$$\operatorname{dom}\Psi := \{x \in X : \Psi(x) \neq \varnothing\},$$
$$\operatorname{range}(\Psi) := \{y \in Y : \text{对某一 } x \in X,\ y \in \Psi(x)\}.$$

即 Ψ 的定义域与值域分别是 Ψ 的图 (graph)

$$\operatorname{gph}(\Psi) := \{(x, y) \in X \times Y : y \in \Psi(x), x \in X\}$$

到 X 与 Y 上的投影. (图) 逆 $\Psi^{-1} : Y \to 2^X$ 定义为

$$\Psi^{-1}(y) := \{x \in X : y \in \Psi(x)\}.$$

由上述定义立即得到 $(\Psi^{-1})^{-1} = \Psi$ 且 $\operatorname{dom}\Psi = \operatorname{range}(\Psi^{-1})$.

多值映射 Ψ 在 $x \in X$ 处被称为是闭的(closed), 若 $x_n \to x$, $y_n \in \Psi(x_n)$, 且 $y_n \to y$, 则 $y \in \Psi(x)$. 称 Ψ 是闭的, 若它在 X 中的每一点均是闭的. 注意到 Ψ 是闭的当且仅当它的图 $\operatorname{gph}(\Psi)$ 是 $X \times Y$ 中的闭子集. 称 Ψ 是凸的, 若它的图 $\operatorname{gph}(\Psi)$ 是 $X \times Y$ 中的凸子集. 或等价地, Ψ 是凸的充分必要条件是对任何 $x_1, x_2 \in X$, $t \in [0, 1]$,

$$t\Psi(x_1) + (1 - t)\Psi(x_2) \subset \Psi(tx_1 + (1 - t)x_2). \tag{2.126}$$

若 Ψ 是凸的, S 是 X 的凸子集, 则 $\Psi(S) := \bigcup_{x \in S} \Psi(x)$ 是 Y 的凸子集. 称 Ψ 是 $\bar{x} \in X$ 处以 $c > 0$ 为模的上 Lipschitz 的 (upper Lipschitzian), 若

$$\Psi(x) \subset \Psi(\bar{x}) + c\|x - \bar{x}\|\overline{B}_Y \tag{2.127}$$

对 $\bar{x}$ 的邻域中的所有 x 均成立.

2.3.1 广义的开映射定理

令 X 与 Y 是 Banach 空间. 泛函分析的一个基本结果是所谓的开映射定理. 被描述为：若 $A : X \to Y$ 是连续的线性算子, 它是映上的, 即 $A(X) = Y$, 则 $0 \in \text{int}\ A(B_X)$. 由于 A 是线性的, 其图是凸集, 由于 A 是连续的, 其图是闭的. A 是映上的条件等价于条件 $0 \in \text{int}A(X)$. 为获得这些性质, 可将开映射定理推广到具有闭凸图的多值函数的情形.

定理 2.70 (广义开映射定理) 设 X 与 Y 是 Banach 空间, $\Psi : X \to 2^Y$ 是闭的凸的多值函数. 令 $y \in \text{int}(\text{range}\ \Psi)$. 则对每一 $x \in \Psi^{-1}(y)$ 及所有的 $r > 0$, 有 $y \in \text{int}\ \Psi(B_X(x, r))$.

该定理的证明基于下述两个引理. 集合 $S \subset X$ 被称为是吸收的 (absorbing), 若任何 $x \in X$, 存在 $t > 0$, 满足 $tx \in S$.

引理 2.71 设 S 是 Banach 空间X 的闭凸子集. 若 S 是吸收的, 则 $0 \in \text{int}(S)$.

证明 由 Baire 引理 2.1 得, 若完备的度量空间是可数闭子集的并集, 则此集合中至少有一在此度量空间中含有内部点的集合.

考虑集合 $S_n := nS$. 由于 S 是吸收的, 有 $X = \bigcup_{n=1}^{\infty} S_n$, 因此由 Baire 引理, 至少有一 S_n 具有非空内部. 于是得到 S 具有非空内部, 即存在 $x \in S$ 及 $\varepsilon > 0$, 满足 $B(x, \varepsilon) \subset S$. 再由于 S 是吸收的, 对某一 $t > 0$, $-tx \in S$. 由凸性得 $B(0, r) \subset S$, 其中 $r = \varepsilon t/(1 + t)$, 因而有 $0 \in \text{int}(S)$. □

定义 2.72 令 S 是向量空间 E 的子集. 称 $x_0 \in S$ 属于 S 的核 (core), 记为 $x_0 \in \text{core}(S)$, 若 $S - \{x_0\}$ 是吸收的. 等价地,

$$\text{core}(S) := \{x \in S : \forall y \in E, \exists \varepsilon > 0, [x - \varepsilon y, x + \varepsilon y] \subset S\}. \tag{2.128}$$

注 2.73 令 S 是 Banach 空间X 的凸子集. 显然, S 的核包含在其内部, 易验证, 若内部非空, 则这两个集合是相等的. 由上述引理, 若凸集 S 还是闭的, 则 $\text{core}(S) = \text{int}(S)$.

用 P_X 与 P_Y 分别记到 X 与 Y 上的投影. 即对点 $(x, y) \in X \times Y$, $P_X(x, y) := x$, $P_Y(x, y) := y$.

引理 2.74 (Robinson 引理) 令 C 是 $X \times Y$ 的闭凸子集. 若 $P_X(C)$ 是有界的, 则 $\text{int}(\text{cl}P_Y(C)) = \text{int}(P_Y(C))$.

证明 只需证 $\text{int}(\text{cl}P_Y(C)) \subset P_Y(C)$. 考虑点 $\bar{y} \in \text{int}(\text{cl}P_Y(C))$. 我们将构造一点 $\bar{x} \in X$ 满足 $(\bar{x}, \bar{y}) \in C$. 这就说明了所需要的包含关系. 令 $\varepsilon > 0$ 满足 $B(\bar{y}, 2\varepsilon) \subset \text{cl}(P_Y(C))$, 令 (x_0, y_0) 是 C 中的一点 (若 C 是空集, 则结论是平凡成立的). 设序列 $(x_k, y_k) \in C$ 由下列形式定义：

当 $(y_k \neq \bar{y})$：

令 $\alpha_k := \varepsilon\|y_k - \bar{y}\|^{-1}$, 则 $w := \bar{y} + \alpha_k(\bar{y} - y_k) \in B(\bar{y}, \varepsilon) \subset \mathrm{cl}(P_Y(C))$.

取 $(u, v) \in C$ 满足 $\|v - w\| \leqslant \frac{1}{2}\|y_k - \bar{y}\|$(因为 $w \in \mathrm{cl}(P_Y(C))$, 这是可能的). 置

$$(x_{k+1}, y_{k+1}) := \frac{\alpha_k}{1+\alpha_k}(x_k, y_k) + \frac{1}{1+\alpha_k}(u, v) \in C.$$

若序列是有限的, 则 $y_k = \bar{y}$, 而 $\bar{x} = x_k$ 即所求的点. 否则, 序列满足

(i) $\|x_{k+1} - x_k\| = \dfrac{\|x_k - u\|}{1+\alpha_k} \leqslant \dfrac{\mathrm{diam}(P_X(C))}{\varepsilon}\|y_k - \bar{y}\|$.

(ii) $\|y_{k+1} - \bar{y}\| = \dfrac{\|v - w\|}{1+\alpha_k} \leqslant \dfrac{1}{2}\|y_k - \bar{y}\|$.

关系式 (ii) 可推出 $\|y_k - \bar{y}\| \leqslant 2^{-k}\|y_0 - \bar{y}\|$. 因此, y_k 收敛到 $\bar{y}$, 结合 (i), 可得序列 $\{x_k\}$ 是 Cauchy 序列. 因为 X 是 Banach 空间, $\{x_k\}$ 有极限 $\bar{x}$. 因为 C 是闭集, 由此得 $(\bar{x}, \bar{y}) \in C$. □

定理 2.70 的证明(Robinson)　不失一般性, 可设 $x = 0, y = 0$ 且 $r = 1$. 考虑集合 $Z := \mathrm{cl}\left(\frac{1}{2}\Psi(B_X)\right)$. 显然, 由于 Ψ 是凸的, Z 是非空闭凸集. 考虑任意点 $y \in Y$. 因为 $0 \in \mathrm{int}(\mathrm{range}\,\Psi)$, 存在 $\alpha > 0$ 满足 $\alpha y \in \mathrm{range}(\Psi)$, 因此, 存在 $x \in X$, 满足 $\alpha y \in \Psi(x)$. 进一步, 对任意的 $t \in (0, 1)$,

$$t\alpha y = t\alpha y + (1-t)0 \in t\Psi(x) + (1-t)\Psi(0) \subset \Psi(tx + (1-t)0) = \Psi(tx),$$

因此, 对某一充分小的 $t > 0$, 有 $t\alpha y \in \Psi\left(\frac{1}{2}B_X\right)$. 因此集合 $\Psi\left(\frac{1}{2}B_X\right)$ 进而集合 Z 是吸收的. 因此, 由引理 2.71 我们得到 0 是 Z 的一内点. 结果, 存在 $\eta > 0$, 满足 $\eta B_Y \subset \mathrm{int}\left(\mathrm{cl}\,\Psi\left(\frac{1}{2}B_X\right)\right)$. 考虑集合 $C := (\mathrm{gph}(\Psi)) \cap \left(\frac{1}{2}\overline{B}_X \times Y\right)$. 显然, $\Psi\left(\frac{1}{2}\overline{B}_X\right) = P_Y(C)$. 进一步, C 是闭的凸的集合且 $P_X(C) \subset \frac{1}{2}\overline{B}_X$ 是有界的. 因此, 由引理 2.74 得 $\mathrm{int}(\mathrm{cl}P_Y(C)) = \mathrm{int}(P_Y(C))$, 从而

$$\begin{aligned}\eta B_Y \subset \mathrm{int}\left(\mathrm{cl}\,\Psi\left(\frac{1}{2}B_X\right)\right) &\subset \mathrm{int}\left(\mathrm{cl}\,\Psi\left(\frac{1}{2}\overline{B}_X\right)\right)\\ = \mathrm{int}\left(\Psi\left(\frac{1}{2}\overline{B}_X\right)\right) &\subset \mathrm{int}\,(\Psi(B_X)),\end{aligned}$$

这证得结论. □

注 2.75　上述证明中的条件 $y \in \mathrm{int}(\mathrm{range}\,\Psi)$ 仅用于推导 $[\mathrm{range}(\Psi) - y]$ 是吸收的. 因此, 这一条件可被似乎更弱的条件 $y \in \mathrm{core}(\mathrm{range}\,\Psi)$ 代替. 然而, 由上述定理的结果可得, 由此可推出 $y \in \mathrm{int}(\mathrm{range}\,\Psi)$, 因此对闭凸的多值函数, 这两个条件是等价的.

2.3.2 开性、稳定性与度量正则性

令 X 与 Y 是 Banach 空间. 引入多值函数 Ψ 的开性的概念.

定义 2.76 称多值函数 $\Psi : X \to 2^Y$ 在 $(x_0, y_0) \in \operatorname{gph}(\Psi)$ 以线性率 $\gamma > 0$ 为开的, 若存在 $t_{\max} > 0$ 及 (x_0, y_0) 的邻域 N 满足对 $\forall (x, y) \in \operatorname{gph}(\Psi) \cap N$, $\forall t \in [0, t_{\max}]$, 下述包含关系成立

$$y + t\gamma B_Y \subset \Psi(x + tB_X). \tag{2.129}$$

命题 2.77 若多值函数 Ψ 是凸的, 则 Ψ 在点 $(x_0, y_0) \in \operatorname{gph}(\Psi)$ 处为开的充分必要条件是存在正数 η, ν, 满足

$$y_0 + \eta B_Y \subset \Psi(x_0 + \nu B_X). \tag{2.130}$$

证明 显然, 取 $\nu = t_{\max}, \eta = \gamma t_{\max}$, 由 (2.129) 可得 (2.130). 相反地, 设 Ψ 是凸的, 且 (2.130) 成立. 不妨设 $x_0 = 0$, $y_0 = 0$. 取

$$N := \nu B_X \times \frac{1}{2}\eta B_Y, \tag{2.131}$$

令 $(x, y) \in \operatorname{gph}(\Psi) \cap N$. 用 $y \in \Psi(x)$, Ψ 的凸性, 及 (2.130), 对任何 $t \in [0, 1]$, 得到

$$\begin{aligned} y + \frac{1}{2}t\eta B_Y &= (1-t)y + t\left(y + \frac{1}{2}\eta B_Y\right) \\ &\subset (1-t)y + t\eta B_y \\ &\subset (1-t)\Psi(x) + t\Psi(\nu B_X) \\ &\subset \Psi((1-t)x + t\nu B_X) \\ &\subset \Psi(x + 2t\nu B_X). \end{aligned}$$

则令

$$\gamma = \frac{\eta}{4\nu}, \quad t_{\max} = 2\nu, \tag{2.132}$$

由 N 定义为 (2.131), 可得 (2.129). □

注 2.78 我们已经证明了下述估计: 若 (2.130) 成立, 则由 (2.131) 与 (2.132) 定义 $N, \gamma, t_{\max}$, 对所有 $(x, y) \in \operatorname{gph}(\Psi) \cap N$, 有 (2.129) 式成立.

注意到, 若多值函数 Ψ 是闭凸的, 则由广义开映射定理2.70, 由正则性条件 $y_0 \in \operatorname{int}(\operatorname{range} \Psi)$ 可得, 存在 η 与 ν 满足 (2.130). 显然, 相反的结论亦成立. 因此得到下述结果.

命题 2.79 设多值函数 $\Psi : X \to 2^Y$ 是闭的凸的. 则 Ψ 在 (x_0, y_0) 处是开的当且仅当 $y_0 \in \operatorname{int}(\operatorname{range} \Psi)$.

与开性密切相联系的一个性质是度量正则性(metric regularity). 回顾 $\operatorname{dist}(x, S) = \inf\limits_{z \in S} \|x - z\|$, 若 S 是空集, 则置 $\operatorname{dist}(x, S) = +\infty$.

定义 2.80 称多值函数 $\Psi: X \to 2^Y$ 在 $(x_0, y_0) \in \mathrm{gph}\Psi$ 以率 c 度量正则的, 若对 (x_0, y_0) 邻域中的所有的 (x, y), 有

$$\mathrm{dist}(x, \Psi^{-1}(y)) \leqslant c\, \mathrm{dist}(y, \Psi(x)). \tag{2.133}$$

度量正则性的概念对估计优化问题可行域的距离是具有指导性的. 下述结果表明, 事实上, 开性与度量正则性的概念是相互等价的.

定理 2.81 多值函数 $\Psi: X \to 2^Y$ 在 $(x_0, y_0) \in \mathrm{gph}\Psi$ 以率 c 为度量正则的当且仅当 Ψ 在 (x_0, y_0) 处以 $\gamma = c^{-1}$ 为率是开的.

证明 设 Ψ 在 (x_0, y_0) 处以 $\gamma > 0$ 为率是开的. 令 $t_{\max} > 0$ 与 N 为定义 2.76 给出的. 不失一般性, 可设 N 具有下述形式

$$N = \varepsilon_x B_X \times \varepsilon_y B_Y. \tag{2.134}$$

若有必要, 减小 $t_{\max}$, 再设

$$t_{\max}\gamma \leqslant \frac{1}{2}\varepsilon_y. \tag{2.135}$$

令 (x, y) 满足

$$\|x - x_0\| < \varepsilon_x^1, \quad \|y - y_0\| < \varepsilon_y^1, \tag{2.136}$$

其中 ε_x^1, ε_y^1 是满足下式的正常数

$$\varepsilon_x^1 \leqslant \varepsilon_x, \quad \gamma\varepsilon_x^1 + \varepsilon_y^1 \leqslant t_{\max}\gamma. \tag{2.137}$$

注意到上述关系表明 $\varepsilon_y^1 \leqslant \frac{1}{2}\varepsilon_y$. 下证关系式 (2.133) 在 $c = \gamma^{-1}$ 时是成立的. 事实上, 由于 $\varepsilon_y^1 \leqslant \varepsilon_y$ 且 Ψ 在 (x_0, y_0) 处是开的, 存在 $x^* \in \Psi^{-1}(y)$, 满足 $\|x^* - x_0\| \leqslant \gamma^{-1}\|y - y_0\|$. 从而有

$$\mathrm{dist}\left(x, \Psi^{-1}(y)\right) \leqslant \|x - x^*\| \leqslant \|x - x_0\| + \gamma^{-1}\|y - y_0\| \leqslant \varepsilon_x^1 + \gamma^{-1}\varepsilon_y^1.$$

结果, 若

$$\mathrm{dist}\left(y, \Psi(x)\right) \geqslant \gamma(\varepsilon_x^1 + \gamma^{-1}\varepsilon_y^1) = \gamma\varepsilon_x^1 + \varepsilon_y^1$$

(尤其, 若 $\Psi(x) = \varnothing$, 这是成立的), 则证得结论. 否则, 由 (2.137), 对充分小 $\alpha > 0$, 存在 $y_\alpha \in \Psi(x)$ 满足

$$\|y - y_\alpha\| \leqslant \mathrm{dist}\left(y, \Psi(x)\right) + \alpha < \gamma\varepsilon_x^1 + \varepsilon_y^1 \leqslant t_{\max}\gamma. \tag{2.138}$$

则, 由 (2.135)~(2.137) 可得

$$\|y_\alpha - y_0\| \leqslant \|y_\alpha - y\| + \|y - y_0\| < t_{\max}\gamma + \varepsilon_y^1 \leqslant \varepsilon_y. \tag{2.139}$$

因此, $(x, y_\alpha) \in \mathrm{gph}(\Psi) \cap N$. 结合 (2.139) 与 Ψ 于 (x_0, y_0) 处的开性可推出, 存在 $x' \in \Psi^{-1}(y)$ 满足 $\|x' - x\| \leqslant \gamma^{-1}\|y - y_\alpha\|$. 于是得到

$$\mathrm{dist}\left(x, \Psi^{-1}(y)\right) \leqslant \|x' - x\| \leqslant \gamma^{-1}\|y - y_\alpha\| \leqslant \gamma^{-1}\,\mathrm{dist}\,(y, \Psi(x)) + \gamma^{-1}\alpha.$$

由于 $\alpha > 0$ 是任意的, 当 $c = \gamma^{-1}$ 时 (2.133) 成立.

相反地, 设 Ψ 在 (x_0, y_0) 以 $c > 0$ 率为度量正则的. 令 $(x, y) \in \mathrm{gph}(\Psi)$, $z \in Y$ 满足 $\|y - z\| < tc^{-1}$. 则对充分接近于 (x_0, y_0) 的 (x, y) 与充分小的 $t > 0$, 有

$$\mathrm{dist}\left(x, \Psi^{-1}(z)\right) \leqslant c\,\mathrm{dist}\,(z, \Psi(x)) \leqslant c\|z - y\| < t.$$

这意味着存在 $w \in \Psi^{-1}(z)$ 满足 $\|w - x\| < t$. 因此有 $z \in \Psi(x + tB_X)$. 证毕. □

注 2.82 后续的证明需要前面证明中涉及的常数的精确估计, 注意到已经证得: 若 N 具有 (2.134) 的形式, 条件 (2.135) 与 (2.136) 成立, 且 $\varepsilon_x^1, \varepsilon_y^1$ 是满足 (2.137) 的正常数, 则不等式 (2.133) 成立.

由上述的注记及作为广义开映射定理2.70 结果的命题 2.79, 由上述定理可得下述结果.

定理 2.83 (Robinson-Ursescu 稳定性定理) *令 $\Psi : X \to 2^Y$ 是闭凸的多值函数. 则 Ψ 在 $(x_0, y_0) \in \mathrm{gph}(\Psi)$ 处度量正则的充分必要条件是正则性条件 $y_0 \in \mathrm{int}(\mathrm{range}\,\Psi)$ 成立. 更精确地, 设 (2.130) 成立, (x, y) 满足*

$$\|x - x_0\| < \frac{1}{2}\nu, \quad \|y - y_0\| < \frac{1}{8}\eta. \tag{2.140}$$

则当 $c = 4\nu/\eta$ 时, (2.133) 成立.

证明 度量正则性与正则性条件 $y_0 \in \mathrm{int}(\mathrm{range}\,\Psi)$ 的等价性是命题 2.79 与定理 2.81 的结果. 因此, 只需要检验常数的估计. 为此, 用注 2.78. 由 (2.131) 与 (2.132) 得到 Ψ 是开的, 其中

$$\varepsilon_x = \nu, \quad \varepsilon_y = \frac{1}{2}\eta, \quad \gamma = \frac{\eta}{4\nu}, \quad t_{\max} = \nu.$$

事实上, (2.131) 与 (2.132) 容许我们取 $t_{\max}$ 等于 2ν, 但将 $t_{\max}$ 减至 ν 可验证 $t_{\max}\gamma = \dfrac{\eta}{4} = \dfrac{1}{2}\varepsilon_y$, 从而 (2.135) 成立. 接着置

$$\varepsilon_x^1 = \frac{1}{2}\nu, \quad \varepsilon_y^1 = \frac{1}{8}\eta,$$

它们满足 (2.137). 则由注 2.82, 当 $c = \gamma^{-1} = 4\nu/\eta$ 时 (2.133) 成立. □

2.3.3 非线性约束系统的稳定性

这一节讨论由下述形式定义的抽象约束的可行域的稳定性

$$\Phi(u) := \{x \in X : G(x,u) \in K\}, \tag{2.141}$$

其中 X 与 Y 是 Banach 空间, U 是拓扑空间, K 是 Y 的闭凸子集, 且 $G : X \times U \to Y$ 是连续映射. 映射 $F : X \to Y$ 称为在集 $S \subset X$ 上以模为 κ Lipschitz 连续的, 若

$$\|F(x_1) - F(x_2)\| \leqslant \kappa \|x_1 - x_2\|, \quad \forall x_1, x_2 \in S.$$

为应用稳定性定理 2.83, 需要下述结果. 先考虑仅依赖于 x 的连续映射$G : X \to Y$, 闭凸集 $K \subset Y$, 与相应的多值函数

$$\mathcal{F}_G(x) := G(x) - K \tag{2.142}$$

是从 X 到 2^Y 的映射. 关系 $y_0 \in \mathcal{F}_G(x_0)$ 意味着 $G(x_0) - y_0 \in K$. 设 $y_0 \in \mathcal{F}_G(x_0)$, 若 $\mathcal{F}_G$ 在 (x_0, y_0) 处是度量正则的, 即

$$\mathrm{dist}(x, \mathcal{F}_G^{-1}(y)) \leqslant c\, \mathrm{dist}(y, \mathcal{F}_G(x)), \tag{2.143}$$

或等价地,

$$\mathrm{dist}(x, G^{-1}(K + y)) \leqslant c\, \mathrm{dist}(G(x) - y, K), \tag{2.144}$$

其中 $c > 0$ 是某一常数, (x, y) 在 (x_0, y_0) 的邻域中 (见定义 2.80). 设 $H : X \to Y$ 是另一连续映射. 下述定理给出条件, 使得相应的多值函数 $\mathcal{F}_H(x) := H(x) - K$ 在 $(x_0, H(x_0) - G(x_0) + y_0)$ 处也是度量正则的.

定理 2.84 *设 $G : X \to Y$ 是连续映射. 设相应的多值函数 $\mathcal{F}_G$ 在 (x_0, y_0) 处以率 $c > 0$ 度量正则, 差值映射 $D(x) := G(x) - H(x)$ 在 x_0 的邻域以模 $\kappa < c^{-1}$ Lipschitz 连续. 则多值函数 $\mathcal{F}_H$ 在 $(x_0, y_0 - D(x_0))$ 处以率 $c(\kappa) := c(1 - c\kappa)^{-1}$ 度量正则, 即*

$$\mathrm{dist}(x, \mathcal{F}_H^{-1}(y)) \leqslant c(\kappa)\, \mathrm{dist}(y, \mathcal{F}_H(x)) \tag{2.145}$$

对充分接近于 $(x_0, y_0 - D(x_0))$ 的 (x, y) 成立.

证明 设 $\eta_x > 0$, $\eta_y > 0$ 满足, 只要

$$\|y - y_0\| < \eta_y \ \text{ 且 } \ \|x - x_0\| < \eta_x, \tag{2.146}$$

则不等式 (2.143) 成立. 要证 (2.145) 对满足

$$\|y - (y_0 - D(x_0))\| < \eta_y^1 \ \text{ 且 } \ \|x - x_0\| < \eta_x^1 \tag{2.147}$$

的 x 与 y 成立, 其中正数 η_x^1 与 η_y^1 是下面将要估计的正常数.

由于度量正则性不受多值函数加上常数项的影响, 不妨设 $D(x_0)=0$. 令 (x,y) 满足 (2.147). 注意到

$$x^* \in \mathcal{F}_H^{-1}(y) \Longleftrightarrow x^* \in \mathcal{F}_G^{-1}(y+D(x^*)). \tag{2.148}$$

令 $\beta \in (c\kappa, 1)$ 与 $\varepsilon > 0$ 满足

$$(1+\varepsilon)c\kappa < \beta. \tag{2.149}$$

我们将给出从 $x_1 = x$ 出发的序列 $\{x_k\}$ 的构造, 它将满足下述递归关系

(i) $x_{k+1} \in \mathcal{F}_G^{-1}(y+D(x_k))$.

(ii) $\|x_k - x_{k+1}\| \leqslant (1+\varepsilon)\operatorname{dist}(x_k, \mathcal{F}_G^{-1}(y+D(x_k)))$. (2.150)

令 η_x^1, η_y^1 满足下述条件 (附加条件之后将给出)

$$\eta_x^1 < \eta_x, \quad \eta_y^1 + \kappa\eta_x^1 < \eta_y. \tag{2.151}$$

则满足 (2.147) 的 (x,y) 满足

$$\|y+D(x)-y_0\| < \eta_y \text{ 且 } \|x-x_0\| < \eta_x,$$

因此, 由 $\mathcal{F}_G$ 在 (x_0,y_0) 的度量正则性得到

$$\operatorname{dist}\left(x, \mathcal{F}_G^{-1}(y+D(x))\right) \leqslant c\operatorname{dist}\left(G(x)-y-D(x),K\right) = c\operatorname{dist}\left(y,\mathcal{F}_H(x)\right).$$

因此, 存在 $x_2 \in \mathcal{F}_G^{-1}(y+D(x))$, 由 (2.149) 得

$$\|x_2 - x_1\| \leqslant c(1+\varepsilon)\operatorname{dist}\left(y,\mathcal{F}_H(x)\right) < \kappa^{-1}\beta\operatorname{dist}\left(y,\mathcal{F}_H(x)\right). \tag{2.152}$$

令

$$\alpha(\eta) := \sup\left\{\|G(x)-G(x_0)\| : x \in B(x_0,\eta)\right\} \tag{2.153}$$

记 G 在 x_0 处的连续性的模数. 则

$$\begin{aligned}\operatorname{dist}\left(y,\mathcal{F}_H(x)\right) &= \operatorname{dist}\left(G(x)-y-D(x),K\right) \\ &\leqslant \|G(x)-G(x_0)\| + \|y-y_0\| + \|D(x)\| \\ &\leqslant \alpha(\eta_x^1) + \kappa\eta_x^1 + \eta_y^1.\end{aligned} \tag{2.154}$$

因此, 由 (2.152),

$$\|x_2 - x_1\| \leqslant \kappa^{-1}\beta(\alpha(\eta_x^1) + \kappa\eta_x^1 + \eta_y^1). \tag{2.155}$$

则对充分小的 $\eta_x^1 > 0$ 与 $\eta_y^1 > 0$, 下述关系对 $k=2$ 成立

$$x_k \in B(x_0,\eta_x) \text{ 且 } y+D(x_k) \in B(y_0,\eta_x). \tag{2.156}$$

现在用数学归纳法证明：对充分小的 $\eta_x^1 > 0$ 与 $\eta_y^1 > 0$, 上述关系对所有的 k 均成立. 若对 $k \geqslant 2$, 这些关系是成立的, 则 $x_k \in \mathcal{F}_G^{-1}(y + D(x_{k-1}))$ 满足

$$\begin{aligned}\operatorname{dist}\left(x_k, \mathcal{F}_G^{-1}(y + D(x_k))\right) &\leqslant c \operatorname{dist}(y + D(x_k), \mathcal{F}_G(x_k)) \\ &\leqslant c\|D(x_k) - D(x_{k-1})\| \leqslant c\kappa\|x_k - x_{k-1}\|.\end{aligned}$$

注意到, 上述的第二个不等式由下式得到

$$x_k \in \mathcal{F}_G^{-1}(y + D(x_{k-1})) \Longleftrightarrow y + D(x_{k-1}) \in \mathcal{F}_G(x_k).$$

现在设 (2.156) 对所有的 $k < k_0$ 成立, $k_0 > 2$ (已知这对 $k_0 = 3$ 是成立的). 由 (2.150(ii)), 对 $2 \leqslant k < k_0$, 有

$$\|x_{k+1} - x_k\| < c^{-1}\kappa^{-1}\beta \operatorname{dist}\left(x_k, \mathcal{F}_G^{-1}(y + D(x_k))\right) \leqslant \beta\|x_k - x_{k-1}\|,$$

从而 $\|x_{k+1} - x_k\| < \beta^{k-1}\|x_2 - x_1\|$. 由 (2.152) 得

$$\|x_{k_0} - x_1\| < (1-\beta)^{-1}\|x_2 - x_1\| \leqslant \kappa^{-1}\beta(1-\beta)^{-1}\operatorname{dist}(y, \mathcal{F}_H(x)), \tag{2.157}$$

从而

$$\begin{aligned}\|x_{k_0} - x_0\| &\leqslant \|x_{k_0} - x_1\| + \|x_1 - x_0\| \\ &< \eta_x^1 + \kappa^{-1}\beta(1-\beta)^{-1}\operatorname{dist}(y, \mathcal{F}_H(x)),\end{aligned} \tag{2.158}$$

再由 (2.147) 有

$$\begin{aligned}\|y + D(x_{k_0}) - y_0\| &\leqslant \|y - y_0\| + \kappa\|x_{k_0} - x_0\| \\ &\leqslant \eta_y^1 + \kappa\eta_x^1 + \beta(1-\beta)^{-1}\operatorname{dist}(y, \mathcal{F}_H(x)).\end{aligned} \tag{2.159}$$

注意到 (2.154), 由数学归纳法证得, 若 $\eta > 0$ 充分小 (不依赖于 k_0), 则 (2.156) 对所有的 k 是成立的.

由于 X 是完备的, 由上述估计可得序列 x_k 存在且收敛到 $B(x_0, \eta)$ 的闭包中的一点 x^*. 进一步, 由于 $D(\cdot)$ 是连续的且 $\mathcal{F}_G^{-1}$ 是闭的, 有 $x^* \in \mathcal{F}_G^{-1}(y + D(x^*))$, 因此有 $x^* \in \mathcal{F}_H^{-1}(y)$.

由 (2.157) 及 $\mathcal{F}_G$ 的度量正则性得, 对充分接近 (x_0, y_0) 的所有的 (x, y),

$$\operatorname{dist}\left(x, \mathcal{F}_H^{-1}(y)\right) \leqslant \|x - x^*\| \leqslant \kappa^{-1}\beta(1-\beta)^{-1}\operatorname{dist}(y, \mathcal{F}_H(x)).$$

因为 β 可取充分接近于 $c\kappa$, 这完成了证明. □

注意到 K 的凸性没有在上述定理的证明中用到.

注 2.85 再一次, 为将来证明时用到, 我们需要出现在 (2.147) 的常数 $\eta_x^1 > 0$ 与 $\eta_y^1 > 0$ 的更加精确的估计. 这些常数应满足 (2.151), 使得 (2.156) 成立. 注意到 (2.151), (2.154), (2.158) 与 (2.159), 由于可取 β 任意接近于 $c\kappa$, 这些条件当下述关系式成立时是成立的

$$\eta_x^1 + c(1-c\kappa)^{-1}[\alpha(\eta_x^1) + \kappa\eta_x^1 + \eta_y^1] < \eta_x, \tag{2.160}$$

$$\eta_y^1 + \kappa\eta_x^1 + c\kappa(1-c\kappa)^{-1}[\alpha(\eta_x^1) + \kappa\eta_x^1 + \eta_y^1] < \eta_y, \tag{2.161}$$

其中 $\alpha(\cdot)$ 由 (2.153) 定义.

现在, 对 $u \in U$, 考虑在 (2.141) 定义的集合 $\Phi(u)$ 及与映射 $G(x,u)$ 相联系的多值函数 $\mathcal{F}_u(x) := G(x,u) - K$. 显然, 点 x 属于 $\Phi(u)$ 当且仅当 $0 \in \mathcal{F}_u(x)$, 因而集合 $\Phi(u)$ 可表示为 $\Phi(u) = \mathcal{F}_u^{-1}(0)$.

设 $G(x,u)$ 关于 x 是可微的且 $D_xG(x,u)$ 是关于 x 与 u 联合变量的 (以算子范数拓扑) 连续函数 (这当然意味着 $G(\cdot,u)$ 是连续可微的, 因而是 Fréchet 可微的). 考虑点 $x_0 \in \Phi(u_0)$. 我们通过 $G(\cdot,u_0)$ 在 x_0 处的线性化函数来近似多值函数 $\mathcal{F}_u$. 即考虑多值函数

$$\mathcal{F}^*(x) = G(x_0,u_0) + D_xG(x_0,u_0)(x-x_0) - K. \tag{2.162}$$

由中值定理有, 差函数

$$G(x,u) - [G(x_0,u_0) + D_xG(x_0,u_0)(x-x_0)]$$

在 x_0 的邻域 N 内是 Lipschitz 连续的, 其相应的 Lipschitz 常数 $\kappa = \kappa(u)$ 满足

$$\kappa \leqslant \sup_{x\in N} \|D_xG(x,u) - D_xG(x_0,u_0)\|.$$

由 $D_xG(x,u)$ 的连续性, N 可取得充分小, u 与 u_0 充分接近, 常数 κ 可以关于 u 一致地任意小. 结合定理 2.84 可推出, 若线性化多值函数 $\mathcal{F}^*$ 在 $(x_0,0)$ 处是度量正则的, 则 $\mathcal{F}_{u_0}$ 在 $(x_0,0)$ 处亦是度量正则的, 相反地, $\mathcal{F}_{u_0}$ 在 $(x_0,0)$ 处的度量正则性可推出 $\mathcal{F}^*$ 的度量正则性. 相同的论断可以应用到 u 接近到 u_0 的 "一致" 的情况, 定理 2.84 类似的形式可类似地证明.

由于线性化多值函数 $\mathcal{F}^*$ 是闭凸的, 定理 2.83 中的稳定性结果可应用到 $\mathcal{F}^*$ 上. 所需要的正则性条件 $0 \in \text{int}(\text{range}\mathcal{F}^*)$ 变为下述形式

$$0 \in \text{int}\left\{G(x_0,u_0) + D_xG(x_0,u_0)X - K\right\}. \tag{2.163}$$

定义 2.86 称 Robinson 约束规范在满足 $G(x_0,u_0) \in K$ 的点 $x_0 \in X$, 关于映射 $G(\cdot,u_0)$ 及集合 K 是成立的, 若上述正则性条件 (2.163) 是成立的.

下述定理是稳定性定理 2.83 与定理 2.84 的 "一致性" 形式的一个结论.

定理 2.87 (稳定性定理) 设 $x_0 \in \Phi(u_0)$ 使 Robinson 约束规范 (2.163) 成立. 则对 (x_0, u_0) 邻域的 (x, u), 成立

$$\text{dist}\,(x, \Phi(u)) = O\,(\text{dist}(G(x,u), K))\,. \tag{2.164}$$

注 2.88 在 $b_0 := G(x_0, u_0)$ 与 $A_0 := D_xG(x_0, u_0)$ 的小的扰动之下, 约束规范 (2.163) 是稳定的. 即若条件 (2.163) 成立, Y 中的向量 b 与 b_0 充分接近, $A \in \mathcal{L}(X, Y)$ 连续线性算子充分接近 A_0, 则

$$0 \in \text{int}\{b + A(X) - K\}.$$

相对于 b 能保持性质 (2.163) 直接由定义即可得到, 关于 A 能保持该性质可由 (2.163) 与相应的多值函数的度量正则性的等价性并结合定理 2.84 得到. 尤其, 由 (2.163) 可得

$$0 \in \text{int}\{G(x,u) + D_xG(x,u)X - K\}$$

对在 (x_0, u_0) 的充分小邻域中的 (x, u) 是成立的.

考虑映射 $G : X \to Y$ 及点 $x_0 \in G^{-1}(K)$. 多值函数 $\mathcal{F}_G(x) := G(x) - K$ 在点 $(x_0, 0)$ 的度量正则性意味着

$$\text{dist}\left(x, G^{-1}(K-y)\right) \leqslant c\ \text{dist}\,(G(x) + y, K) \tag{2.165}$$

对 $(x_0, 0)$ 某一邻域中所有的 (x, y) 成立, 其中 $c > 0$ 是常数. 若条件 (2.165) 成立, 即多值函数 $\mathcal{F}_G$ 在 $(x_0, 0)$ 处是度量正则的, 称映射 G 在 x_0 处关于 K 是度量正则的. 由定理 2.87 可得, 若 G 是连续可微的且 Robinson 约束规范

$$0 \in \text{int}\{G(x_0) + DG(x_0)X - K\} \tag{2.166}$$

成立, 则 G 在 x_0 处关于 K 是度量正则的. 相反地, 条件 (2.165) 可推出, 对接近于 0 的 y, $\text{dist}(x_0, G^{-1}(K-y)) < +\infty$, 因此集合 $G^{-1}(K-y)$ 非空. 即

$$0 \in \text{int}\{G(X) - K\}. \tag{2.167}$$

若多值函数 $x \mapsto G(x) - K$ 是凸的. 这一条件与度量正则性等价, 这可由定理 2.83 得到.

由定理 2.84 得, 若映射 G 在 x_0 处是度量正则的, 则其在 x_0 处线性化映射也是度量正则的. 对线性化映射, 条件 (2.167) 取 (2.166) 的形式. 因此, 我们得到下述结论 (与定理 2.83 相比较).

命题 2.89 连续可微映射 $G : X \to Y$ 在点 $x_0 \in G^{-1}(K)$ 处关于集 K 是度量正则的充分必要条件是 Robinson 约束规范 (2.166) 成立.

在非线性分析中, 定理 2.87 是一个强有力的工具. 我们看一些例子. 设空间 U 是 Banach 空间, 映射 $G(x,u)$ 在 (x_0,u_0) 的邻域内是连续可微的. 考虑集合

$$L(u) := \{x \in X : G(x_0,u_0) + DG(x_0,u_0)(x-x_0,u-u_0) \in K\}, \tag{2.168}$$

它通过 $G(x,u)$ 的线性化得到. 设 Robinson 约束规范(2.163) 成立, 则 (2.164) 成立. 令 $x \in L(u)$. 由 (2.164), 存在 $c>0$, 当 (x,u) 充分接近 (x_0,u_0) 时, 有

$$\operatorname{dist}\Big(x,\Phi(u)\Big) \leqslant c\,\|G(x,u) - G(x_0,u_0) - DG(x_0,u_0)(x-x_0,u-u_0)\|.$$

由于 $G(x,u)$ 是连续可微的, 有

$$\lim_{\substack{(x,u)\to(x_0,u_0)\\ x\in L(u)}} \frac{\operatorname{dist}(x,\Phi(u))}{\|x-x_0\|+\|u-u_0\|} = 0. \tag{2.169}$$

以同样的方式, 用线性化多值函数的度量正则性可以证明

$$\lim_{\substack{(x,u)\to(x_0,u_0)\\ x\in \Phi(u)}} \frac{\operatorname{dist}(x,L(u))}{\|x-x_0\|+\|u-u_0\|} = 0. \tag{2.170}$$

即 $\Phi(u)$ 与 $L(u)$ 在 (x_0,u_0) 处是一阶等价的. 尤其, 在 (2.169) 与 (2.170) 中取 $u=u_0$, 可得下述结果.

命题 2.90 设映射 $G: X\to Y$ 在点 $x_0 \in \Phi := G^{-1}(K)$ 处是连续可微的, 且 Robinson 约束规范(2.166) 成立. 则

$$\operatorname{dist}(x,\Phi) = o(\|x-x_0\|), \quad 对\ x - x_0 \in \mathcal{T}, \tag{2.171}$$

$$\operatorname{dist}(h,\mathcal{T}) = o(\|h\|), \quad 对\ x_0 + h \in \Phi, \tag{2.172}$$

其中 $\mathcal{T} := \{h : G(x_0) + DG(x_0)h \in K\}$.

推论 2.91 设映射 $G: X\to Y$ 在点 $x_0 \in \Phi := G^{-1}(K)$ 处是连续可微的, 且 Robinson 约束规范(2.166) 成立. 则 Φ 在 x_0 处的内切锥、余切锥、Clarke 切锥是重合的, 且

$$T_\Phi(x_0) = \{h \in X : DG(x_0)h \in T_K(G(x_0))\}. \tag{2.173}$$

证明 由命题 2.90 可得, 内切锥$T_\Phi^i(x_0)$ 与余切锥$T_\Phi(x_0)$ 相等, 且等式 (2.173) 成立.

由命题 2.56 得 Clarke 切锥$T_\Phi^c(x_0)$ 包含于 $T_\Phi(x_0)$. 因此, 只需证明 $T_\Phi(x_0) \subset T_\Phi^c(x_0)$. 考虑 $h \in T_\Phi(x_0)$. 令 $x_n \to x_0$, $x_n \in \Phi\setminus\{x_0\}$, $t_n \downarrow 0$. 有 $y_n := G(x_n) \in K$, 由于 G 是连续的, 有 $y_n \to y_0$, 其中 $y_0 := G(x_0)$. 进一步, 由于 G 是连续

可微的, 有 $G(x_n+t_nh)=y_n+t_nDG(x_0)h+o(t_n)$. 由于集合 K 是凸的, 由命题 2.57 得 $T_K(y_0)=T_K^c(y_0)$. 因为 $DG(x_0)h\in T_K(y_0)$, 存在序列 $z_n\in K$ 满足 $\|z_n-(y_n+t_nDG(x_0)h)\|=o(t_n)$. 从而

$$\operatorname{dist}(G(x_n+t_nh),K)\leqslant\|z_n-G(x_n+t_nh)\|=o(t_n),$$

由 (2.164) 得 $\operatorname{dist}(x_n+t_nh,\Phi)=o(t_n)$, 即存在 $w_n\in\Phi$ 满足 $(w_n-x_n)/t_n\to h$. 这表明 $h\in T_\Phi^c(x_0)$, 即得要证之结论. □

例 2.92 设 $\Phi(u)$ 是下述形式由有限个约束定义的集合

$$\Phi(u)=\{x:g_i(x,u)=0,i=1,\cdots,q;\ g_i(x,u)\leqslant 0,i=q+1,\cdots,p\}. \tag{2.174}$$

通过取 $K=\{0\}\times\mathbb{R}_-^{p-q}$, $G(x,u)=(g_1(x,u),\cdots,g_p(x,u)):X\times U\to\mathbb{R}^p$. 在 $\mathbb{R}^p$ 中赋予 ℓ_1 范数, 即 $\|y\|:=\sum_{i=1}^{p}|y_i|$ ($\mathbb{R}^p$ 中所有范数是等价的). 对这样的 K, 点 $y=(y_1,\cdots,y_p)\in\mathbb{R}^p$ 及 ℓ_1 范数, 可得

$$\operatorname{dist}(y,K)=\sum_{i=1}^{q}|y_i|+\sum_{i=q+1}^{p}[y_i]_+,$$

其中 $[a]_+:=\max\{a,0\}$. 所以, 此种情形, 估计 (2.164) 具有下述形式

$$\operatorname{dist}(x,\Phi(u))=O\left(\sum_{i=1}^{q}|g_i(x,u)|+\sum_{i=q+1}^{p}[g_i(x,u)]_+\right). \tag{2.175}$$

例 2.93 设 Y 是由 $p\times p$ 对称矩阵构成的线性空间 $\mathcal{S}^p$, $K:=\mathcal{S}_-^p$ 是负半定矩阵构成的锥 (半定规划的例子). 考虑 $\mathcal{S}^p$ 中的下述范数: 对 $A\in\mathcal{S}^p$, 定义 $\|A\|:=\max\limits_{1\leqslant i\leqslant p}|\lambda_i(A)|$, 其中 $\lambda_i(A)$ 是矩阵 A 的特征值. 此种情况

$$\operatorname{dist}(A,\mathcal{S}_-^p)=[\lambda_{\max}(A)]_+,$$

(2.164) 具有下述形式

$$\operatorname{dist}(x,\Phi(u))=O([\lambda_{\max}(G(x,u))]_+). \tag{2.176}$$

例 2.94 设 $Y=C(\Omega)$, $K=C_-(\Omega)$ 是非正值函数 $y\in C(\Omega)$ 的集合. 则对任何 $y\in C(\Omega)$,

$$\operatorname{dist}(y,K)=\max_{\omega\in\Omega}[y(\omega)]_+,$$

从而 (2.164) 具有下述形式

$$\operatorname{dist}(x,\Phi(u))=O\left(\max_{\omega\in\Omega}[G(x,u)(\omega)]_+\right). \tag{2.177}$$

2.3.4 约束规范条件

设 X 与 Y 是 Banach 空间, $G: X \to Y$ 是连续可微映射, K 是 Y 的闭凸子集. $\mathcal{R}_K(y)$ 记 K 在 $y \in Y$ 处的雷达方向锥, $\mathcal{R}_K(y)$ 的拓扑包是切锥 $T_K(y)$(见 (2.83) 与 (2.91)).

这一节讨论具体情况下与 Robinson 约束规范

$$0 \in \operatorname{int}\{G(x_0) + DG(x_0)X - K\} \tag{2.178}$$

等价的各种正则性条件.

命题 2.95 设 $G(x_0) \in K$. 则下述的 (2.179) 与 (2.180) 互相等价且与 Robinson 约束规范(2.178) 等价

$$0 \in \operatorname{core}\{G(x_0) + DG(x_0)X - K\}, \tag{2.179}$$

$$DG(x_0)X - \mathcal{R}_K(G(x_0)) = Y. \tag{2.180}$$

证明 设条件 (2.180) 成立. 考虑如下定义的多值函数 $\mathcal{M}: X \times \mathbb{R} \to 2^Y$,

$$\mathcal{M}(x,t) := \begin{cases} DG(x_0)x - t(K - G(x_0)), & 若\ t \geqslant 0, \\ \varnothing, & 若\ t < 0. \end{cases}$$

显然, $\mathcal{M}$ 是闭的凸的多值函数, 且

$$\operatorname{range}(\mathcal{M}) = DG(x_0)X - \mathcal{R}_K(G(x_0)).$$

结果, 由 (2.180) 得到 $0 \in \operatorname{int}(\operatorname{range}\mathcal{M})$. 由广义开映射定理2.70 得 $0 \in \operatorname{int}(\mathcal{M}(X \times [0,1]))$, 即得到 (2.178). 关系 (2.178)⇒(2.179) 是显然的, 由于 $K - G(x_0) \subset \mathcal{R}_K(G(x_0))$ 及 (2.180) 的左端是锥, 可得关系 (2.179)⇒(2.180). □

若 K 是锥, 则 $\mathcal{R}_K(G(x_0)) = K + [\![G(x_0)]\!]$, 从而在此种情况下, 条件 (2.180) 具有下述形式

$$DG(x_0)X - K + [\![G(x_0)]\!] = Y. \tag{2.181}$$

一个很自然的问题是, 在条件 (2.180) 中是否可以将雷达锥$\mathcal{R}_K(G(x_0))$ 用相应的切锥 $T_K(G(x_0))$ 来代替. 然而, 这样的条件不与 (2.178) 或 (2.179) 等价.

例 2.96 考虑 $Y = L_2[0,1]$, 锥 $K \subset L_2[0,1]$ 由几乎处处非正值函数组成 (与例 2.64 比较). 设 $DG(x_0) = 0$, $y_0 = G(x_0)$ 满足 $y_0(t) < 0$ 对几乎处处的 $t \in [0,1]$ 成立. 则 $T_K(y_0) = Y$, 而 $\mathcal{R}_K(y_0) \neq Y$. 本例的雷达锥与切锥具有这样的性质是因为锥 K 生成整个空间 $Y = L_2[0,1]$, 即 $K - K = Y$, 而 K 在 Y 中不具有内部点.

在某些情况之下, 涉及关于 K 的切锥的条件与 Robinson 约束规范条件是等价的, 如下述命题所示.

命题 2.97　考虑条件

$$DG(x_0)X - T_K(G(x_0)) = Y, \tag{2.182}$$

$$[DG(x_0)X]^{\perp} \cap N_K(G(x_0)) = \{0\}, \tag{2.183}$$

$$\text{cl}\,[DG(x_0)X - T_K(G(x_0))] = Y, \tag{2.184}$$

$$\text{cl}\,[DG(x_0)X - \mathcal{R}_K(G(x_0))] = Y. \tag{2.185}$$

上述条件及 Robinson 约束规范(2.178) 满足下述关系

$$(2.178) \Rightarrow (2.182) \Rightarrow (2.183) \Leftrightarrow (2.184) \Leftrightarrow (2.185).$$

若还有, 集合 $\mathcal{A} := DG(x_0)X - \mathcal{R}_K(G(x_0))$ 具有非空的相对内部, 则条件 (2.182)~(2.185) 是相互等价的, 且与 Robinson 约束规范(2.178) 等价.

证明　(2.178) $\Rightarrow$ (2.182). 由命题 2.95, (2.178) 等价于 (2.180). 因为 $\mathcal{R}_K(G(x_0)) \subset T_K(G(x_0))$, 由 (2.180) 推出 (2.182).

(2.182) $\Rightarrow$ (2.183). 对 (2.182) 两边计算极锥, 在右端得到 $Y^- = \{0\}$, 由于 $[T_K(G(x_0))]^- = N_K(G(x_0))$, 用 (2.31), 在左端得到 $[DG(x_0)X]^{\perp} \cap N_K(G(x_0))$. 结果, 由 (2.182) 可得到 (2.183).

(2.183) $\Leftrightarrow$ (2.184) $\Leftrightarrow$ (2.185). 在 (2.183) 两边取极锥, 用 (2.32) 可得第一等价关系, 注意到 $\mathcal{R}_K(G(x_0))$ 的闭包是 $T_K(G(x_0))$, 可得第二等价关系.

关系 (2.185) 意味着 $\text{cl}(\mathcal{A}) = Y$. 由定理 2.17, 若 $\mathcal{A}$ 具有非空相对内部, 则 (2.185) 等价于 $\mathcal{A} = Y$, 即由命题 2.95, (2.185) 与 (2.178) 等价. 证毕. □

注意到, 条件 (2.180), (2.182), (2.184) 与 (2.185) 左端中的减号可替换为加号. 同时, 若 $DG(x_0)$ 是映上的, 则上述条件 (2.178) 与 (2.182)~(2.185) 成立.

推论 2.98　设 Y 是有限维的或 K 具有非空内部. 则 $\mathcal{A} := DG(x_0)R - \mathcal{R}_K(G(x_0))$ 具有非空的相对内部, 因而条件 (2.182)~(2.185) 是等价的, 且等价于 Robinson 约束规范(2.178).

Robinson 约束规范(2.178) 涵盖许多特殊情况. 例如, K 是单点集的情况, 即等式约束的情形, (2.178) 简化为条件 $DG(x_0): X \to Y$ 是映上的. 这里有一些其他有意思的情况.

引理 2.99　若 K 具有非空内部, 则 (2.178) 等价于

$$\exists h \in X : G(x_0) + DG(x_0)h \in \text{int}(K). \tag{2.186}$$

证明　若存在 $h \in X$ 满足 $y := G(x_0) + DG(x_0)h$ 属于 $\text{int}(K)$, 则存在 $\varepsilon > 0$, $B(y, \varepsilon) \subset K$. 因此有 $B(0, \varepsilon) \subset G(x_0) + DG(x_0)h - K$, 这证得 (2.178).

相反地, 设凸集 $G(x_0)+DG(x_0)X$ 与 $\text{int}(K)$ 具有空集的交. 由于 K 具有非空的内部, 由第一分离定理2.13 得, 存在非零 $\lambda\in Y^*$, 满足对所有的 $h\in X$ 与 $k\in K$, 有 $\langle\lambda, G(x_0)+DG(x_0)h\rangle \geqslant \langle\lambda, k\rangle$. 令 $y\in Y$ 满足 $\langle\lambda, y\rangle<0$. 则对 $t>0$, ty 不属于 $G(x_0)+DG(x_0)X-K$①, 可推出后者不包含以 0 为中心的球, 因此 (2.178) 不成立. □

上述引理的刻画 (2.186) 可解释为线性系统的 Slater 条件 (见 2.3.5 节).

在某些情况下, 约束由集合的卡氏积的形式给出, 即 Y 是两个 Banach 空间Y_1 与 Y_2 的卡氏积, $K=K_1\times K_2\subset Y_1\times Y_2$, 其中 K_1 与 K_2 分别是 Y_1 与 Y_2 的闭凸子集. 则 $G(x)=(G_1(x), G_2(x))$, $G_i(x)\in Y_i$, $i=1,2$. 我们用 $[DG(x_0)]^{-1}$ 记多值函数, 它的图是 $DG(x_0)$ 的图的逆, 即

$$[DG(x_0)]^{-1}(y):=\{h\in X: DG(x_0)h=y\}.$$

引理 2.100 设约束由上述的卡氏积形式给出, 设 $DG_1(x_0)$ 是映上的, 则 (2.178) 等价于

$$0\in\text{int}\left\{G_2(x_0)+DG_2(x_0)[DG_1(x_0)^{-1}(K_1-G_1(x_0))]-K_2\right\}. \tag{2.187}$$

证明 条件 (2.178) 意味着存在 $\varepsilon_1>0$, $\varepsilon_2>0$ 满足若 $y_i\in Y_i, \|y_i\|\leqslant\varepsilon_i$, $i=1,2$, 则存在 $h\in X$, 满足对 $i=1,2$, 有 $G_i(x_0)+DG_i(x_0)h-y_i\in K_i$. 由于对 $i=1$ 时, 这一条件等价于 $h\in DG_1(x_0)^{-1}(y_1+K_1-G_1(x_0))$, 一个等价条件为

$$y_2\in G_2(x_0)+DG_2(x_0)[DG_1(x_0)^{-1}(y_1+K_1-G_1(x_0))]-K_2.$$

取 $y_1=0$, 得到 (2.187) 是 (2.178) 的必要条件.

相反地, 设 (2.187) 成立. 则由开映射定理 (推论 2.29), 存在 $M>0$ 满足对每一 $y_1\in Y_1$, 存在 $h_1\in X$ 满足 $G_1(x_0)h_1=y_1$ 且 $\|h_1\|\leqslant M\|y_1\|$. 因此, $h\in X$ 满足 $y\in G(x_0)+DG(x_0)h-K$ 当且仅当 $h_2:=h-h_1$ 满足 $DG_1(x_0)h_2\in K_1-G_1(x_0)$ 且

$$G_2(x_0)+DG_2(x_0)h_2-K_2\ni y_2-DG_2(x_0)h_1.$$

给定 $\varepsilon>0$, 由于 $\|h_1\|\leqslant M\|y_1\|$, 取 ε_1 与 ε_2 充分小, 得到右端属于 $B(0,\varepsilon)$. 因此, 由 (2.187) 可推出这样的 h 是存在的, 即 (2.178) 成立②. □

在 K_2 具有非空内部的情形, 用引理 2.99 证明中相同的分离定理, 可以得到下述结果.

推论 2.101 设 $DG_1(x_0)$ 是映上的, K_2 具有非空的内部. 则 (2.178) 等价于存在 $h\in X$, 满足

---

① 原著中为 $G(x_0)+DG(x_0)-K$.

② 原著中是 (2.187) 成立.

$$
\begin{aligned}
&G_1(x_0) + DG_1(x_0)h \in K_1, \\
&G_2(x_0) + DG_2(x_0)h \in \operatorname{int}(K_2).
\end{aligned} \tag{2.188}
$$

若 $K_1 = \{0\}$, 即第一约束是等式型的约束, 则条件 $DG_1(x_0)$ 是映上的是 (2.178) 的必要条件. 结果, 此种情况, (2.178) 等价于

$$
\begin{aligned}
&DG_1(x_0) \text{ 是映上的}, \\
&0 \in \operatorname{int}\left\{G_2(x_0) + DG_2(x_0)[\ker DG_1(x_0)] - K_2\right\}.
\end{aligned} \tag{2.189}
$$

若 $K_1 = \{0\}$, K_2 具有非空内部, 则 (2.178) 等价于

$$
\begin{aligned}
&DG_1(x_0) \text{ 是映上的}, \\
&\exists\, h \in \ker DG_1(x_0) \text{ 满足 } G_2(x_0) + DG_2(x_0)h \in \operatorname{int}(K_2).
\end{aligned} \tag{2.190}
$$

尤其, 若 Φ 由有限多个等式和不等式约束定义, 如在 (2.174) 中那样, 则 (2.178) 等价于 Mangasarian-Fromovitz 约束规范:

$$
\begin{aligned}
&DG_i(x_0)\ i = 1, \cdots, q \text{ 是线性无关的}, \\
&\exists\, h \in X : Dg_i(x_0)h = 0,\ i = 1, \cdots, q, \quad Dg_i(x_0)h < 0,\ i \in I(x_0),
\end{aligned} \tag{2.191}
$$

其中 $I(x_0)$ 记 x_0 处起作用的不等式约束的指标集.

再回到卡氏积形式的一般约束, 由引理 2.100, 若 $Y_1 = X$, $G_1(x) = x$, $\forall x \in X$, 则下述条件等价于 Robinson 约束规范(2.178):

$$
0 \in \operatorname{int}\left\{G_2(x_0) + DG_2(x_0)(K_1 - x_0) - K_2\right\}. \tag{2.192}
$$

改变记号, 再来表述这一结果是合适的. 设可行集具有下述形式

$$
\Phi := \{x \in X : x \in Q, G(x) \in K\}, \tag{2.193}
$$

其中 $Q \subset X$ 与 $K \subset Y$ 是闭凸集. 通过置映射 $x \mapsto (x, G(x))$, 凸集 $Q \times K$ 是空间 $X \times Y$ 的子集, 则这可视为 (2.141) 形式约束的特殊形式. 令 $x_0 \in \Phi$ 是可行点. 此种情况, 正则性条件 (2.178) 与 (2.180) 分别等价于

$$
0 \in \operatorname{int}\left\{G(x_0) + DG(x_0)(Q - x_0) - K\right\} \tag{2.194}
$$

与

$$
DG(x_0)(\mathcal{R}_Q(x_0)) - \mathcal{R}_K(G(x_0)) = Y \tag{2.195}
$$

(它们也是彼此等价的).

若 Φ 由 (2.193) 定义, K 具有非空内部, 则 (2.178) 等价于

$$\exists\, x' \in Q : G(x_0) + DG(x_0)(x' - x_0) \in \operatorname{int}(K). \tag{2.196}$$

例 2.102 设 Ω 是非空的紧致的 Hausdorff 拓扑空间, $Y := C(\Omega)$, $K := C_-(\Omega)$ (见例 2.63). 令 $G : X \to Y$ 是连续可微映射, $G(x_0) \in K$. 考虑对应函数 $g(x, \omega) := G(x)(\omega)$. 则约束 $G(x) \in K$ 可表示为形式 $g(x, \omega) \leqslant 0$, $\forall \omega \in \Omega$. 此例中集 (锥)$K$ 具有非空内部, 因此 Robinson 约束规范(2.178) 等价于 (2.186). 用 $\Delta(x_0)$ 表示在 x_0 处起作用的约束集, 即

$$\Delta(x_0) := \{\omega \in \Omega : g(x_0, \omega) = 0\}.$$

因为 Ω 是紧致的, 不难证得 (2.186) 成立当且仅当

$$\exists\, h \in X, \forall\, \omega \in \Delta(x_0) : D_x g(x_0, \omega) h < 0. \tag{2.197}$$

将正则性条件 (2.197) 视为增广的 Mangasarian-Fromovitz 约束规范. 在此例中, 它等价于 Robinson 约束规范.

2.3.5 凸映射

现在讨论凸映射的概念, 它可视为凸函数的概念的推广. 对这样的映射, 约束规范取具体的形式, 在不同的应用中是有用途的.

定义 2.103 *称映射 $G : X \to Y$ 关于闭凸集 $C \subset Y$ 是凸的, 或简言之, G 是 C 凸的, 若相应的多值函数 $\mathcal{M}_G := G(x) + C$ 是凸的*[①]

回顾凸多值函数的定义 (见 2.3 节) 得到, G 是 C 凸的当且仅当对任意的 x_1, $x_2 \in X$ 与 $t \in [0, 1]$,

$$tG(x_1) + (1 - t)G(x_2) - G(tx_1 + (1 - t)x_2) + C \subset C. \tag{2.198}$$

一个很重要的结论是, 包含关系 (2.198) 成立当且仅当

$$tG(x_1) + (1 - t)G(x_2) - G(tx_1 + (1 - t)x_2)$$

属于集合 C 的回收锥C^∞(关于回收锥的讨论, 见 2.1.4 节). 这意味着 G 是 C 凸的当且仅当它关于 C 的回收锥是凸的. 由于回收锥C^∞ 是闭的、凸的, 由对偶关系 $((C^\infty)^-)^- = C^\infty$, (2.198) 等价于对任何 $\lambda \in (C^\infty)^-$, 有

$$\langle \lambda, tG(x_1) + (1 - t)G(x_2) - G(tx_1 + (1 - t)x_2) \rangle \leqslant 0.$$

这等价于函数 $\langle -\lambda, G(\cdot) \rangle$ 的凸性, 对任意的 $\lambda \in (C^\infty)^-$ 由上面的讨论得 G 的 C 凸性是 C 的回收锥的性质. 若 C_1 与 C_2 是 X 的两个闭凸子集, $C_1^\infty \subset C_2^\infty$, G 是关于 C_1 为凸的, 则 G 关于 C_2 是凸的. 尤其, 若 G 是 C 凸的, 则对任意 $x \in C$, G 是

① 应为 $\mathcal{M}_G(x)$.

$T_C(x)$ 凸的. 若 G 是线性的, 则它关于任何闭凸集 C 均是凸的. 相反地, 若 C 的回收锥是 $\{0\}$, 则 G 的 C 凸性可推出其线性性. 在无穷维空间的情形, 即使 C 是无界的, C 的回收锥也可能是 $\{0\}$(见例 2.43).

尤其, 若 C 是闭凸锥, 则 (2.198) 成立当且仅当

$$tG(x_1) + (1-t)G(x_2) - G(tx_1 + (1-t)x_2) \in C. \tag{2.199}$$

这等价于条件

$$G(tx_1 + (1-t)x_2) \preceq_C tG(x_1) + (1-t)G(x_2), \tag{2.200}$$

其中 $\preceq_C$ 记由锥 C 生成的序关系, 即 $a \preceq_C b$ 意味着 $b - a \in C$. 例如, 若 $Y = \mathbb{R}^p$ 且 $C = \mathbb{R}^p_+$, 则映射 $G(x) = (g_1(x), \cdots, g_p(x))$ 是 C 凸的当且仅当坐标函数 $g_i : X \to \mathbb{R}$ 是通常意义下凸的. 还有, 条件 (2.199) 与 (2.200) 等价于函数 $\langle -\lambda, G(\cdot)\rangle$ 的凸性, 对任何 $\lambda \in C^-$.

现在考虑约束系统 (2.193), 其中 Q 与 K 分别是 X 与 Y 中的闭凸子集. 设映射 $G(\cdot)$ 关于集合 $C := -K$ 是凸的, 即 $\mathcal{F}_G(x) := G(x) - K$ 是凸的. 若映射 G 是连续的, 则多值函数 $\mathcal{F}_G(x)$ 是闭的. 进一步, 考虑多值函数

$$\mathcal{F}^*_{G,Q}(x) := \begin{cases} \mathcal{F}_G(x), & \text{若 } x \in Q, \\ \varnothing, & \text{若 } x \notin Q, \end{cases} \tag{2.201}$$

它亦是凸的、闭的. 凸多值函数的正则性条件, 即 $0 \in \text{int}(\text{range}\mathcal{F}^*_{G,Q})$ 具有下述形式: 对 Y 中 0 的邻域中所有的 y, 存在 $x \in Q$ 满足 $G(x) + y \in K$, 可以表示为

$$0 \in \text{int}\{G(Q) - K\}. \tag{2.202}$$

命题 2.104　*若映射 $G(x)$ 是 $(-K)$ 凸的, 且是连续可微的, 则条件 (2.202) 等价于在每一可行点 $x_0 \in \Phi$ 处的 Robinson 约束规范.*

证明　由定理 2.83, 条件 (2.202) 等价于度量正则性, 结果由命题 2.89 得, 等价于 Robinson 约束规范. □

定义 2.105　*称对系统 (2.193), (广义) Slater 条件成立, 若存在 $\bar{x} \in Q$ 满足 $G(\bar{x}) \in \text{int}(K)$.*

显然, Slater 条件可推出 (2.202), 若集合 K 具有非空内部, 相反的结论亦成立.

命题 2.106　*设 K 是凸集, 具有非空内部, $G(\cdot)$ 关于 $(-K)$ 是凸的. 则 (2.202) 等价于 Slater 条件.*

证明　显然, Slater 条件可推出 (2.202). 相反地, 假设 Slater 条件不成立, 即 $0 \notin G(Q) - \text{int}(K)$. 我们给出论断: $G(Q) - \text{int}(K)$ 是凸集. 显然, 这一集合是开的. 由

第一分离定理 2.13, 存在 $y^* \in Y^*\backslash\{0\}$ 满足 $\langle y^*, y\rangle \geqslant 0$ 对所有的 $y \in G(Q) - \mathrm{int}(K)$. 因此, 对所有的 $y \in G(Q) - K$, $\langle y^*, y\rangle \geqslant 0$, 这同 (2.202) 矛盾.

剩下的是要证明论断. 由于 G 关于 $(-K)$ 是凸的, 它关于锥 $(-K)^\infty$ 亦是凸的, 因此集合 $G(Q) - K^\infty$ 是凸的. 进一步, 若 $k \in K^\infty$, 则 $k + \mathrm{int}(K) \subset \mathrm{int}(K)$, 因此 (由于凸集之差是凸的) 集合 $G(Q) - \mathrm{int}(K)$ 等于集合 $(G(Q) - K^\infty) - \mathrm{int}(K)$, 是凸集. □

2.4 凸 函 数

这一节给出凸函数理论的某些基本结果：凸函数连续性的充分条件、凸函数的共轭性质以及次微分性质. 这一节的结果将应用于 2.5 节的对偶理论中.

除非特别说明, 这一节设 X 与 X^* 是 (局部凸的拓扑向量) 成对空间 (见定义 2.26). 若 X 是 Banach 空间, X^* 是其对偶, 则标准的成对空间 X 与 X^* 满足 X(或 X^*) 赋予强 (或弱 *) 拓扑. 若 X 是自反的 Banach 空间, X 与 X^* 均被赋予强拓扑, 则 X 与 X^* 是成对的空间.

2.4.1 连续性

这一节给出正常凸函数在其定义域内部上连续性的一些充分条件. 注意到, 有可能正常的凸函数具有稠密的定义域而其内部是空集. 例如, 函数 $f : L_1[0,1] \to \overline{\mathbb{R}}$ 定义为

$$f(x) = \begin{cases} 0, & x \in L_2[0,1], \\ +\infty, & \text{否则}. \end{cases}$$

命题 2.107 设 $f : X \to \overline{\mathbb{R}}$ 是正常凸函数. 则下述结论是等价的：

(i) f 在点 x_0 的某邻域上是上有界的.

(ii) 存在 $x_0 \in \mathrm{dom}f$ 满足 f 在 x_0 处是连续的.

(iii) $\mathrm{int}(\mathrm{epi}f) \neq \varnothing$.

(iv) $\mathrm{int}(\mathrm{epi}f) \neq \varnothing$ 且 f 在 $\mathrm{int}(\mathrm{dom}f)$ 上是连续的.

若还有 X 是赋予强拓扑的 Banach 空间, 则条件 (i)~(iv) 等价于下述条件：

(v) $\mathrm{int}(\mathrm{dom}f) \neq \varnothing$ 且 f 在 $\mathrm{int}(\mathrm{dom}f)$ 上是局部 Lipschitz 连续的.

证明 设条件 (i) 成立, 即存在 x_0 的邻域 N 及常数 c 满足：对所有 $x \in N$ 满足 $f(x) \leqslant c$. 不妨设 $x_0 = 0$, $f(x_0) = 0$, $c > 0$. 由 f 的凸性, 对任意 $x \in N$, $t \in [0,1]$,

$$f(tx) = f(tx + (1-t)0) \leqslant tf(x) + (1-t)f(0) = tf(x) \leqslant tc. \tag{2.203}$$

因此, 对任意的 $\varepsilon > 0$, 对所有的 $x \in \varepsilon c^{-1}N$ 有 $f(x) \leqslant \varepsilon$. 进一步, 存在 $\gamma > 0$, 若 $x \in (-\gamma)N$, 则 $-\gamma^{-1}x \in N$, 由 f 的凸性得

$$0 = f(0) = f\left(\frac{1}{1+\gamma}x + \frac{\gamma}{1+\gamma}\left(-\frac{x}{\gamma}\right)\right) \leqslant \frac{1}{1+\gamma}f(x) + \frac{\gamma}{1+\gamma}c. \tag{2.204}$$

则对所有的 $x \in (-\gamma)N$ 有 $f(x) \geqslant -\gamma c$, 因此对充分小的 $\gamma > 0$, $f(x) \geqslant -\varepsilon$. 由于 ε 是任意的, 因而表明 f 在 0 处是连续的.

显然, 若 f 在 x_0 处是连续的, 则它在 x_0 的邻域 N 上是有界的, 因此存在 $c \in \mathbb{R}$ 满足 $N \times (c, +\infty) \subset \mathrm{epi}f$. 结果, (ii)⇒ (iii).

设条件 (iii) 成立, 存在 X 的一开子集 V 与 $c \in \mathbb{R}$ 满足 $N \times (c, +\infty) \subset \mathrm{epi}f$. 则 $V \subset \mathrm{dom}f$, 且 f 在 V 上是上有界的. 因此, (iii) 可推出 (i), 从而 (i)~(iii) 是等价的.

显然, 由 (iv) 可推出 (iii). 所以, 为证明 (iv) 与 (i)~(iii) 的等价性, 只需要证明, 若 $\mathrm{int}(\mathrm{epi}f)$ 是非空的, 则

$$P_X(\mathrm{int}(\mathrm{epi}f)) = \mathrm{int}(\mathrm{dom}\, f). \tag{2.205}$$

令 (x_0, c) 是 $\mathrm{epi}f$ 的内部点, $U \subset \mathrm{epi}f$ 是它的开邻域. 考虑点 $x \in \mathrm{int}(\mathrm{dom}\, f)$. 则存在点 $\bar{x} \in \mathrm{dom}\, f$ 及某一 $\alpha \in (0,1)$ 满足 $\alpha x_0 + (1-\alpha)\bar{x} = x$. 从而有 $\alpha U + (1-\alpha)(\bar{x}, f(\bar{x})) \subset \mathrm{epi}f$ 是开集, 此集合到 X 上的投影是 x 的邻域. 这表明 (2.205) 之右端包含在它的左端. 相反的包含关系是显然的, 因此 (2.205) 成立.

现在设 X 是 Banach 空间. (v) 推出 (i) 是显然的. 通过验证相反的结论, 我们完成证明. 考虑点 $x_0 \in \mathrm{int}(\mathrm{dom}\, f)$. 则 f 在 x_0 处是连续的, 因而存在 $c > 0$ 与 $r > 0$, 满足对所有的 $x \in B(x_0, r)$, 有 $|f(x)| \leqslant c$. 令 x_1, $x_2 \in B(x_0, r/4)$. 则

$$f(x_2) - f(x_1) = f(x_1 + tv) - f(x_1) \leqslant 2tc,$$

其中 $v := (r/2)\|x_2 - x_1\|^{-1}(x_2 - x_1)$, $t = (2/r)\|x_2 - x_1\|$. 由于 $B(x_1, r/2) \subset B(x_0, r)$, $\|v\| = r/2$, $t \in [0,1]$ 及

$$\begin{aligned} f(x_1 + tv) &= f((1-t)x_1 + t(v + x_1)) \leqslant (1-t)f(x_1) + tf(v + x_1) \\ &\leqslant f(x_1) + t(|f(x_1)| + f(v + x_1)) \leqslant f(x_1) + 2tc. \end{aligned}$$

得到 f 在 $B(x_0, r/4)$ 上是 Lipschitz 连续的. 因为有 (2.205), 条件 (v) 成立. □

下述命题表明, 有限值点处邻域的凸性与一致上有界性可推出函数的正常性与连续性. 注意到, 在本命题中, 一致有界性的条件是本质的 (见例 2.114).

命题 2.108　*设 $f : X \to \overline{\mathbb{R}}$ 是凸函数, 存在点 $x_0 \in X$, 值 $f(x_0)$ 是有限的, $f(x)$ 在 x_0 的邻域上是上有界的. 则 f 是正常的且在 x_0 处连续.*

证明　设 N 是 x_0 的邻域, c 是常数, 满足对任何 $x \in N$ 均有 $f(x) \leqslant c$. 不失一般性, 设 $x_0 = 0$, $f(x_0) = 0$ 且 $c > 0$. 则对任何 $x \in X$, 存在 $\gamma > 0$ 满足 $x \in (-\gamma)N$. 由 (2.204) 得 $f(x) \geqslant -\gamma c > -\infty$, 因此 f 是正常的. 由命题 2.107 可得 f 在 x_0 处的连续性. □

若空间 X 是有限维的, 由任何点 $x \in \text{int}(\text{dom}\, f)$ 有邻域 $N \subset \text{dom}\, f$, 它是 $\text{dom}\, f$ 有限点的凸包. 结果, f 在 N 上是上方有界的. 得到下述结果.

推论 2.109 *设 f 是定义在有限维空间 X 上的凸函数. 则 f 在 $\text{int}(\text{dom}\, f)$ 上是连续的.*

函数 $f: X \to \overline{\mathbb{R}}$ 的凸包(convex hull), 记为 $\text{conv}f$, 定义为不超过 f 的最大凸函数, 或等价地, 为其上图包含 $\text{conv}(\text{epi}f)$ 的最大的凸函数, 即

$$(\text{conv}f)(x) = \inf\{\alpha : (x, \alpha) \in \text{conv}(\text{epi}f)\}. \tag{2.206}$$

函数 $f: X \to \overline{\mathbb{R}}$ 的下半连续包, 记为 $\text{lsc}f$, 定义为上图为 f 的上图的拓扑包的函数, 函数 f 是下半连续的充分必要条件是 $\text{lsc}f = f$. 处理函数 f 的闭包 (记为 $\text{cl}f$) 也是适宜的, 它定义为

$$\text{cl}f(\cdot) := \begin{cases} \text{lsc}f(\cdot), & \text{若对所有的} x \in X, \text{lsc}f(x) > -\infty, \\ -\infty, & \text{若至少存在一} x \in X, \text{lsc}f(x) = -\infty. \end{cases}$$

称 f 是闭的(closed), 若 $\text{cl}f = f$. 由上述定义得, 正常凸函数是闭的当且仅当它是下半连续的. 由凸集的拓扑包是凸的, 有若函数 f 是凸的, 则 $\text{lsc}f$ 与 $\text{cl}f$ 也是凸的.

命题 2.110 *设 $f: X \to \overline{\mathbb{R}}$ 是凸函数, 若 $\text{lsc}f$ 在某一点处取有限值, 则 $\text{lsc}f$ 与 f 是正常的. 否则,*

$$\text{lsc}f(x) = \begin{cases} -\infty, & \text{若 } x \in \text{cl}(\text{dom}\, f), \\ +\infty, & \text{若 } x \notin \text{cl}(\text{dom}\, f). \end{cases} \tag{2.207}$$

证明 若 $\text{lsc}f$ 在某一点处取有限值, 因为它是凸的, 也是 l.s.c., 由引理 2.33 得, 它是正常的. 关系式 $f(\cdot) \geqslant \text{lsc}f(\cdot)$ 及至少存在一点 x 满足 $f(x) < +\infty$, 这表明 f 亦是正常的.

若 $\text{lsc}f$ 在其定义域上取值 $-\infty$, 因为 $\text{lsc}f$ 的定义域包含 f 的定义域, 由 $\text{lsc}f$ 的下半连续性, 可得 (2.207) 的第一关系是成立的. 第二关系是显然的. □

若凸函数f 的定义域至少有两个不同的点, 则对所有的 $x \in X$, 有

$$f(x) \geqslant \liminf_{x' \to x} f(x'). \tag{2.208}$$

事实上, 若 $f(x_1)$ 与 $f(x_2)$ 是有限的, 且 $x_1 \neq x_2$, 由凸性 (例 2.17), 对任意 $\varepsilon > 0$, 存在充分小的 $t > 0$ 满足 $f(tx_1 + (1-t)x_2) \leqslant f(x_2) + \varepsilon$. 这可推出在 $x = x_2$ 处的 (2.208). 若 $f(x_1)$ 与 $f(x_2)$ 均小于 $+\infty$ 且其一为 $-\infty$, 这也是显然的. 因此, 对凸函数, 其定义域至少有两个不同的点, 有

$$\text{lsc}f(x) = \liminf_{x' \to x} f(x'), \quad \forall x \in X. \tag{2.209}$$

在下述结论中, 移掉上方有界性的假设, 取而代之用函数的下半连续性.

命题 2.111　设 f 是定义在 Banach 空间X 上的凸的下半连续函数, 至少在某一点处取有限值. 则 f 在 $\mathrm{int}(\mathrm{dom}\, f)$ 上是正常的连续的.

证明　令 $x_0 \in \mathrm{int}(\mathrm{dom}\, f)$, $-\infty < f(x_0) < \alpha < +\infty$, $S := \{x \in X : f(x) \leqslant \alpha\}$ 是相应的水平集. 因 f 是凸的, 则 S 是凸集, 因为 f 是 l.s.c., 则 S 是闭的. 进一步, 集合 $S - x_0$ 是吸收的. 这是由于对任何 $h \in X$, 由命题 2.107, 一维函数 $\phi(t) := f(x_0 + th)$ 在 $\mathbb{R}$ 的零的邻域内是连续的. 由引理 2.71 得 $x_0 \in \mathrm{int}(S)$, 因此 $f(x)$ 在 x_0 的邻域上是上方有界的. 由命题 2.107 可证得结果. □

尤其, 由上述命题可知, 若 f 是定义在 Banach 空间X 上的 l.s.c. 凸的实值函数 (即 $-\infty < f(x) < +\infty$ 对所有的 $x \in X$ 成立), 则 f 在 X 上是连续的. 这一命题有一个很有意思的结论. 令 $\{f_i\}_{i \in I}$ 是一类下半连续的凸函数. 考虑最大值函数 $f(x) := \sup_{i \in I} f_i(x)$. 则函数 f 也是 l.s.c. 和凸的, 因而在其定义域的内部是连续的. 尤其, 若 f 是实值的, 则 f 在 X 上是连续的.

2.4.2　共轭性

设 X 与 X^* 是成对空间, 令 $f : X \to \overline{\mathbb{R}}$ 是增广实值 (可能非凸的) 函数. 如下定义的函数 $f^* : X^* \to \overline{\mathbb{R}}$,

$$f^*(x^*) := \sup_{x \in X} \{\langle x^*, x\rangle - f(x)\} \tag{2.210}$$

称为 f 的共轭函数(conjugate function). 类似地, 对 $g : X^* \to \overline{\mathbb{R}}$, 它的共轭 $g^* : X \to \overline{\mathbb{R}}$ 定义为

$$g^*(x) := \sup_{x^* \in X^*} \{\langle x^*, x\rangle - g(x^*)\}. \tag{2.211}$$

尤其, 可考虑函数 $f^{**} = (f^*)^*$,

$$f^{**}(x) := \sup_{x^* \in X^*} \{\langle x^*, x\rangle - f^*(x^*)\}. \tag{2.212}$$

对任何 $x \in X$ 与 $x^* \in X^*$, 下述的 Young-Fenchel 不等式 (Young-Fenchel inequality) 成立, 它由定义 (2.210) 立即可以得到

$$f(x) \geqslant \langle x^*, x\rangle - f^*(x^*). \tag{2.213}$$

进一步, 对所有的 $x \in X$, Young-Fenchel 不等式可推出 $f(x) \geqslant f^{**}(x)$. 因为 f^* 的上图是半空间

$$\{(x^*, c) \in X^* \times \mathbb{R} : c \geqslant \langle x^*, x\rangle - f(x)\}$$

的交, 所以 f^* 是凸的下半连续的. 类似地, f^{**} 是凸的下半连续的.

命题 2.112 *若 $f: X \to \overline{\mathbb{R}}$ 是正常的下半连续的凸函数. 则 f^* 亦是正常的凸函数.*

证明 因为 f^* 是共轭函数, 且是凸的. 因为 f 是正常的, 则对所有的 $x \in X$, $f(x) > -\infty$ 且存在点 x_0, $f(x_0)$ 是有限的. 则对任意 $x^* \in X^*$,

$$f^*(x^*) \geqslant \langle x^*, x_0 \rangle - f(x_0) > -\infty.$$

因为 f 是下半连续的凸的, 它的上图是 $X \times \mathbb{R}$ 的闭凸子集, 有 epi$f$ 可表示为 $X \times \mathbb{R}$ 的一族闭半空间的交. 由于 f 是正常的, 这些闭半空间中至少有一者必是非竖直的. 即存在 $\bar{x}^* \in X^*$, $c \in \mathbb{R}$ 满足: 对所有的 $x \in X$, $f(x) \geqslant \langle \bar{x}^*, x \rangle - c$, 则

$$f^*(\bar{x}^*) \leqslant \sup_{x \in X} \{\langle \bar{x}^*, x \rangle - (\langle \bar{x}^*, x \rangle - c)\} = c,$$

因此 $\bar{x}^* \in \operatorname{dom} f^*$[①]. 这证得 f^* 是正常的. □

下述是关于共轭函数的基本对偶性定理.

定理 2.113 (Fenchel-Moreau-Rockafellar) *设 $f: X \to \overline{\mathbb{R}}$ 是增广实值函数. 则*

$$f^{**} = \mathrm{cl}(\mathrm{conv} f). \tag{2.214}$$

证明 由定义 (2.210) 有, $(x^*, \beta) \in \mathrm{epi} f^*$ 当且仅当

$$f(x) \geqslant \langle x^*, x \rangle - \beta, \quad \forall x \in X. \tag{2.215}$$

换言之, $(x^*, \beta) \in \mathrm{epi} f^*$ 当且仅当 f 的上图被包含在闭的非竖直的半空间中,

$$\{(x, \alpha) \in X \times \mathbb{R} : \alpha \geqslant \langle x^*, x \rangle - \beta\}.$$

于是得到, f^{**} 的上图由闭的非竖直半空间族

$$\{(x, \alpha) \in X \times \mathbb{R} : \alpha \geqslant \langle x^*, x \rangle - \beta\}, \quad (x^*, \beta) \in \mathrm{epi}(f^*)$$

的交给出, 或换言之, f^{**} 是不超过 f 的连续仿射函数的上确界. 因此 epi(f^{**}) 是包含 epi(f) 的非竖直半空间的交. 另一方面, lsc(conv$f$) 的上图是包含 epi(f) 的 (可能是竖直的) 半空间的交, 因为它的上图是包含 epi(f) 的最小闭凸子集.

若对至少一点 $x \in X$ 有 $\mathrm{lsc}(\mathrm{conv} f)(x) = -\infty$, 则由定义有 $\mathrm{cl}(\mathrm{conv} f)(\cdot) = -\infty$, 没有不超过 f 的连续仿射函数存在, 因而 (2.214) 成立. 若 $f(\cdot) = +\infty$, 则 f^{**} 与 cl(conv$f$) 等于 f. 只剩下考虑 lsc(conv$f$) 是正常的, 因而等于 cl(conv$f$) 的情形. 由上面的讨论, 只需证明若 lsc(conv$f$) 是正常的, 点 (x_0, α_0) 不属于 lsc(conv$f$) 的上图, 则该点可以被非竖直的半空间与上图分离开来.

① 原著中为 $\operatorname{dom} f$.

设 $\text{lsc}(\text{conv}f)$ 是正常的. 则 f 至少有不超过它的仿射函数存在. 事实上, 闭凸集 $\text{epi}[\text{lsc}(\text{conv}f)]$ 由包含它的半空间的交给出. 若这些半空间均是竖直的, 则 $\text{lsc}(\text{conv}f)$ 在其定义域上取值 $-\infty$. 因为 $\text{lsc}(\text{conv}f)$ 是正常的, 它的定义域非空, 这与假设 $\text{lsc}(\text{conv}f)(\cdot) > -\infty$ 矛盾. 因此, 存在 $(x^*, \beta) \in X^* \times \mathbb{R}$ 满足

$$\alpha \geqslant \langle x^*, x \rangle - \beta, \quad \forall (x, \alpha) \in \text{epi}[\text{lsc}(\text{conv}f)]. \tag{2.216}$$

若 (x_0, α_0) 不属于 $\text{lsc}(\text{conv}f)$ 的上图, 则存在线性泛函 $(\bar{x}^*, a)$, 它强分离 (x_0, α_0) 与 $\text{lsc}(\text{conv}f)$. 若 $a \neq 0$, 则对应的半空间不是竖直的. 因此设 $a = 0$, 则存在 $\varepsilon > 0$,

$$\langle \bar{x}^*, x_0 \rangle \geqslant \langle \bar{x}^*, x \rangle + \varepsilon, \quad \forall (x, \alpha) \in \text{epi}[\text{lsc}(\text{conv}f)].$$

上述不等式乘以 $\gamma > 0$, 加到不等式 (2.216) 上, 得到

$$\alpha \geqslant \langle \hat{x}^*, x \rangle - \hat{\beta}, \quad \forall (x, \alpha) \in \text{epi}[\text{lsc}(\text{conv}f)], \tag{2.217}$$

其中 $\hat{x}^* := x^* + \gamma \bar{x}^*$, $\hat{\beta} := \beta + \gamma \langle \bar{x}^*, x \rangle - \gamma \varepsilon$. 当 $\varepsilon > 0$ 时, 对充分大的 $\gamma > 0$, 有不等式

$$\alpha_0 < \langle \hat{x}^*, x_0 \rangle - \hat{\beta} \tag{2.218}$$

成立. 因为右端恰好是 $\langle x^*, x_0 \rangle - \beta - \gamma \varepsilon$, 结果得到一非竖直的半空间, 它分离了 (x_0, α_0) 与 $\text{lsc}(\text{conv}f)$ 的上图, 这就完成了证明. □

上述定理表明, $f = f^{**}$ 当且仅当 f 是凸的且是闭的. 尤其, 若对所有的 $x \in X$ 有 $f(x) > -\infty$, 则 $f = f^{**}$ 当且仅当 f 是凸的且是下半连续的.

例 2.114　设 X 是无穷维的局部凸的拓扑向量空间. 令 $\{e_i\}, i \in I$ 是 X 中的代数基, 即每一 $x \in X$ 可唯一地表示为线性组合 $x = \sum\limits_{i \in I} a_i e_i$, 其中只有有限个非零的 $a_i \in \mathbb{R}$, 定义 $f(x) := \sum\limits_{i \in I} a_i$. 函数 f 是线性的 (因而是凸的) 实值函数. 这一函数是非连续的, 因而在 X 中的任何开集上是无界的 (上方无界且下方无界). 结果, $f^*(\cdot) \equiv +\infty$ 且 $f^{**}(\cdot) \equiv -\infty$. 因此, 在此例中, 共轭函数$f^*$ 与 f^{**} 是非正常的, 尽管对每一 $x \in X$, $f(x)$ 是有限值的.

例 2.115　对集合 $S \subset X$, 考虑指示函数(见例 2.67) 与支撑函数

$$\sigma(x^*, S) := \sup_{x \in S} \langle x^*, x \rangle. \tag{2.219}$$

注意到, 空集的上确界被置为 $-\infty$, 则对所有的 x^*, $\sigma(x^*, S) = -\infty$, 或者对所有 x^*, 有 $\sigma(x^*, S) > -\infty$. 于是

$$I_S^*(x^*) = \sup_{x \in X} \{\langle x^*, x \rangle - I_S(x)\} = \sup_{x \in S} \langle x^*, x \rangle = \sigma(x^*, S), \tag{2.220}$$

即指示函数$I_S(\cdot)$ 的共轭是对应的支撑函数$\sigma(\cdot, S)$. 观察到, 指示函数 $I_S(\cdot)$ 是凸的当且仅当 S 是凸的, 是正常的当且仅当 $S \neq \varnothing$, 是 l.s.c. 的当且仅当 S 是闭集. 因此, 由 Fenchel-Moreau-Rockafellar 对偶定理 2.113 得, 若集合 S 是闭的凸的, 则 $\sigma^*(\cdot, S) = I_S(\cdot)$.

支撑函数有下面有用的性质.

命题 2.116 *支撑函数具有下述性质:*

(i) 设 $S_1, S_2 \subset X$ 是凸闭集. 则

$$\sigma(\cdot, S_1) + \sigma(\cdot, S_2) = \sigma(\cdot, S_1 + S_2), \tag{2.221}$$

且 $S_1 \subset S_2$ 当且仅当

$$\sigma(x^*, S_1) \leqslant \sigma(x^*, S_2), \quad \forall x^* \in X^*. \tag{2.222}$$

(ii) 设 $S_i, i \in I$ 是 X 中的一类非空子集. 则

$$\sup_{i\in I}\{\sigma(\cdot, S_i)\} = \sigma\left(\cdot, \mathrm{cl}\left(\mathrm{conv}\left(\bigcup_{i\in I} S_i\right)\right)\right). \tag{2.223}$$

(iii) 设 $S \subset X$ 是非空闭凸集合, S^∞ 是它的回收锥, 且 $Q := \mathrm{dom}\,\sigma(\cdot, S) = \{x^* \in X^* : \sigma(x^*, S) < +\infty\}$. 则 $Q^- := S^\infty$.

证明 可得

$$\begin{aligned}\sigma(x^*, S_1 + S_2) &= \sup_{x\in S_1+S_2} \langle x^*, x\rangle \\ &= \sup_{x_1\in S_1,\, x_2\in S_2} \{\langle x^*, x_1\rangle + \langle x^*, x_2\rangle\} \\ &= \sup_{x_1\in S_1} \langle x^*, x_1\rangle + \sup_{x_2\in S_2} \langle x^*, x_2\rangle,\end{aligned}$$

因此 (2.221) 成立.

当 $S_1 \subset S_2$ 时, 由定义立即得到 $\sigma(\cdot, S_1) \leqslant \sigma(\cdot, S_2)$. 相反地, 若 $\sigma(\cdot, S_1) \leqslant \sigma(\cdot, S_2)$, 则

$$I_{S_1}(\cdot) = \sigma^*(\cdot, S_1) \geqslant \sigma^*(\cdot, S_2) = I_{S_2}(\cdot),$$

因此有 $S_1 \subset S_2$. 即 (2.222) 成立.

关系式 (2.223) 由

$$\sup_{i\in I}\{\sigma(x^*, S_i)\} = \sup_{i\in I}\sup_{x\in S_i} \langle x^*, x\rangle = \sup_{x\in \bigcup_{i\in I} S_i} \langle x^*, x\rangle$$

及某一集合与它的闭凸包的支撑函数是相同的这一事实得到.

为证 (iii), 设 $\bar{x} \in S^\infty$, 取 $x \in S$. 则对任意的 $t \geqslant 0$, 有 $x + t\bar{x} \in S$. 因此任何 $x^* \in Q$, 对 $t \geqslant 0$ 满足

$$\langle x^*, x + t\bar{x} \rangle = \langle x^*, x \rangle + t\langle x^*, \bar{x} \rangle \leqslant \sigma(x^*, S) < +\infty,$$

从而有 $\langle x^*, \bar{x} \rangle \leqslant 0$. 结果 $\bar{x} \in Q^-$, 这表明 $S^\infty \subset Q^-$.

相反地, 设 $\bar{x} \in Q^-$. 为证明 $\bar{x} \in S^\infty$, 需要验证对任何 $x \in S$ 有 $x + \bar{x} \in S$. 假设这是不成立的, 即存在 $x \in S$ 有 $x + \bar{x} \notin S$. 则由第二分离定理 2.14, 存在 $x^* \in X^*$ 及 $\varepsilon > 0$ 满足

$$\langle x^*, x + \bar{x} \rangle \geqslant \langle x^*, x' \rangle + \varepsilon, \quad \forall x' \in S. \tag{2.224}$$

由 (2.224) 得 $\sigma(x^*, S) \leqslant \langle x^*, x + \bar{x} \rangle$, 因此有 $x^* \in Q$. 另一方面, 在 (2.224) 中取 $x' = x$, 得到 $\langle x^*, \bar{x} \rangle \geqslant \varepsilon > 0$, 这与假设 $\bar{x} \in Q^-$ 矛盾. 这推出 $Q^- \subset S^\infty$, 因而完成证明. □

尤其, 由 (2.222) 可得, 若 S_1 与 S_2 是闭凸集合, 则

$$\sigma(\cdot, S_1) = \sigma(\cdot, S_2) \text{ 当且仅当 } S_1 = S_2. \tag{2.225}$$

2.4.3 次可微性

泛函 $x^* \in X^*$ 称为 (可能非凸的) 函数 $f: X \to \overline{\mathbb{R}}$ 在点 x 处的次梯度(subgradient), 若 $f(x)$ 是有限的, 且

$$f(y) - f(x) \geqslant \langle x^*, y - x \rangle, \quad \forall y \in X.v \tag{2.226}$$

这表明, 线性函数 $\ell(y) := \langle x^*, y - x \rangle + f(x)$ 是 f 的上图在 $(x, f(x))$ 的支撑超平面(supporting hyperplane), 即 $\ell(x) = f(x)$ 且对所有 $y \in X$ 有 $f(y) \geqslant \ell(y)$.

函数 f 在 x 处所有次梯度的集合称为 f 在 x 处的次微分 (subdifferential), 即

$$\partial f(x) := \{x^* \in X^* : f(y) - f(x) \geqslant \langle x^*, y - x \rangle, \quad \forall y \in X\}. \tag{2.227}$$

作为闭半空间的交, $\partial f(x)$ 是凸集, $\partial f(x)$ 在 X^* 的相容拓扑下是闭的. 函数 f 称为在 x 处为次可微的 (subdifferentiable), 若 $f(x)$ 是有限的且 $\partial f(x) \neq \varnothing$. 次可微函数是正常的, 这是显然的. 然而, 即使空间 X 是有限维的, 并非每一正常凸函数在定义域的每一点处均是次可微的. 例如, 例 2.128 定义的凸函数. 这一凸函数在 $x = 0$ 处不是次可微的.

由 (2.227) 得, $\bar{x}$ 是 (可能为非凸的) 函数 $f(x)$ 在 X 上极小点当且仅当 $f(\bar{x})$ 是有限的且 $0 \in \partial f(\bar{x})$.

注 2.117 $x^* \in \partial f(\bar{x})$ 当且仅当对所有的 $(y,c) \in \text{epi}(f)$, 下述不等式成立:

$$\langle x^*, y-x\rangle - (c - f(x)) \leqslant 0,$$

或等价地, $(x^*,-1)$ 是 $\text{epi}(f)$ 在点 $(x, f(x))$ 处的法向量. 这就得到

$$\partial f(x) \times \{-1\} = N_{\text{epi}(f)}(x, f(x)) \cap (X^* \times \{-1\}). \tag{2.228}$$

注意到不等式 (2.226) 可以表示为下述形式

$$\langle x^*, x\rangle - f(x) \geqslant \langle x^*, y\rangle - f(y), \quad \forall y \in X.$$

注意到共轭函数的定义 (2.210), 这意味着, 上面不等式的右端对 $y \in X$ 取上确界在 $y = x$ 处达到. 因此有

$$x^* \in \partial f(x) \Longleftrightarrow f(x) + f^*(x^*) = \langle x^*, x\rangle. \tag{2.229}$$

也注意到, (2.229) 的左端可推出 $f(x)$ 是有限的 (否则, 由定义 $\partial f(x)$ 是空集), 而其右端可推出 $f(x)$ 与 $f^*(x^*)$ 均是有限的.

命题 2.118 设 $f: X \to \overline{\mathbb{R}}$ 是 (可能非凸的) 函数. 则下述各款是成立的:

(i) 若对某一 $x \in X$, $f^{**}(x)$ 是有限的, 则

$$\partial f^{**}(x) = \underset{x^* \in X^*}{\text{argmax}}\{\langle x^*, x\rangle - f^*(x^*)\}. \tag{2.230}$$

(ii) 若 f 在 x 处是次可微的, 则 $f^{**}(x) = f(x)$.

(iii) 若 $f^{**}(x) = f(x)$ 且是有限的, 则 $\partial f(x) = \partial f^{**}(x)$.

证明 将 (2.229) 应用到 f^{**}, 有 $x^* \in \partial f^{**}(x)$ 当且仅当

$$f^{**}(x) = \langle x^*, x\rangle - f^{***}(x^*).$$

由 Fenchel-Moreau-Rockafellar 定理(定理 2.113) 得 $f^{***} = f^*$, 因而上述等式等价于

$$f^{**}(x) = \langle x^*, x\rangle - f^*(x^*). \tag{2.231}$$

因为由 (2.212), $f^{**}(x)$ 等于 (2.231) 之右端在 $x^* \in X^*$ 取最大值的最大值点集合, 我们得到 (2.230).

若存在 $x^* \in \partial f(x)$, 则由 (2.229) 得 $f(x) \leqslant f^{**}(x)$. 因为 $f(x) \geqslant f^{**}(x)$, 性质 (ii) 得证.

为证明 (iii), 由 (2.229) 得 $x^* \in \partial f(x)$ 当且仅当 $f(x) = \langle x^*, x\rangle - f^*(x^*)$, $x^* \in \partial f^{**}(x)$ 当且仅当 (2.231) 成立, 这证得结论. □

由 (2.229) 及上述命题, 若 x 使得 $f(x)$ 是有限的, 则下述结论等价：

(i) $x^* \in \partial f(x)$.

(ii) $x \in \partial f^*(x^*)$. (2.232)

(iii) $f(x) + f^*(x^*) = \langle x^*, x\rangle$.

回顾由命题 2.118(ii), 存在 $x^* \in \partial f(x)$, 表明 $f(x) = f^{**}(x)$.

这一节的其余部分限于讨论凸函数.

引理 2.119　设 $f : X \to \overline{\mathbb{R}}$ 是凸函数, 它在凸集 S 上取常的有限值. 则 $\partial f(\cdot)$ 在 $\mathrm{ri}(S)$ 上是常值的. 更确切地, 对每一 $x \in S$, $x_0 \in \mathrm{ri}(S)$, 有 $\partial f(x_0) \subset \partial f(x)$.

证明　令 $x \in S$, $x_0 \in \mathrm{ri}(S)$, $x^* \in \partial f(x_0)$, 则对充分小的 $\varepsilon > 0$, $x_0 \pm \varepsilon(x - x_0) \in S$, 则

$$0 = f(x_0 \pm \varepsilon(x - x_0)) - f(x_0) \geqslant \pm\varepsilon\langle x^*, x - x_0\rangle.$$

结果 $\langle x^*, x - x_0\rangle = 0$, 从而对任何 $x' \in X$, $a := f(S)$, 有

$$\begin{aligned} f(x') \geqslant f(x_0) + \langle x^*, x' - x_0\rangle &= a + \langle x^*, x' - x\rangle + \langle x^*, x - x_0\rangle \\ &= f(x) + \langle x^*, x' - x\rangle. \end{aligned}$$

这意味着 $x^* \in \partial f(x)$, 这证得结论. □

注 2.120　上述的证明仅用到这样的事实, 即若 $x \in S$, 对 $\varepsilon > 0$ 充分小, 有 $x_0 \pm \varepsilon(x - x_0) \in S$. 所以, 只需要设 x_0 属于 S 相对于空间 $\mathrm{Sp}\,(S)$ 的核即可.

命题 2.121　设 K 是 X^* 的非空凸子集, 它以 X^* 的相容拓扑为闭的, 令 $f(x) := \sigma(x, K)$ 是相应的支撑函数. 若 $x \in \mathrm{dom}\, f$, 则

$$\partial f(x) := \mathrm{argmax}\{\langle x^*, x\rangle : x^* \in K\}. \tag{2.233}$$

证明　支撑函数 $f(\cdot)$ 是正常的下半连续凸函数, 得 $f^*(\cdot) = I_K(\cdot)$(见 (2.220)). 得到 $f^{**}(x) = f(x)$, 因而有 $\partial f^{**}(x) = \partial f(x)$. 由公式 (2.230) 即得公式 (2.233). □

例 2.122　设 Ω 是非空的紧致的 Hausdorff 拓扑空间, $X := C(\Omega)$(见例 2.37). $C(\Omega)$ 的对偶空间是 Ω 上的 (有限符号的) 正则 Borel 测度. 考虑最大值函数

$$\psi(x) := \sup_{\omega\in\Omega} x(\omega).$$

显然, $\psi : X \to \mathbb{R}$ 是实值的正齐次连续的 (事实上, 甚至是模 1 Lipschitz 连续的) 凸函数. 不难证明, ψ 是 Ω 上的概率测度(probability measures) 集 $\mathcal{P}_\Omega$ 的支撑函数, 即 $\psi(x) = \sigma(x, \mathcal{P}_\Omega)$, 其中

$$\mathcal{P}_\Omega := \left\{\mu \in C(\Omega)^* : \int_\Omega \mathrm{d}\mu = 1, \mu \succeq 0\right\}, \tag{2.234}$$

$\mu \succeq 0$ 记非负的测度 μ. 事实上, 若 $\mu \in \mathcal{P}_\Omega$ 且 $x \in C(\Omega)$, 则

$$\int_\Omega x(\omega)\mathrm{d}\mu(\omega) \leqslant \psi(x)\mu(\Omega) = \psi(x). \tag{2.235}$$

相反地, 令 $x \in C(\Omega)$. 因为 Ω 是紧致的闭的, 集合 $\operatorname*{argmax}_{\omega\in\Omega} x(\omega)$ 是非空的. 考虑被集合 $\operatorname*{argmax}_{\omega\in\Omega} x(\omega)$ 支撑的测度 $\mu \in \mathcal{P}_\Omega$(这样的测度是存在的, 如取 Dirac 测度 $\mu = \delta(\bar{\omega})$, 其中 $\bar{\omega} \in \operatorname*{argmax}_{\omega\in\Omega} x(\omega)$). 对这样的测度 μ, (2.235) 取等式, 这证得论断.

集合 $\mathcal{P}_\Omega$, 因为是两个弱 * 闭凸集, 即 $[C_-(\Omega)]^-$ 和 $\left\{\mu \in C(\Omega)^* : \int_\Omega \mathrm{d}\mu = 1\right\}$ 的交集, 它是弱 * 闭的凸的. 由于它还是有界的, 得到 $\mathcal{P}_\Omega$ 在 X^* 中是凸的弱 * 紧致的. 结果由命题 2.121 得, $\partial\psi(x)$ 与 $\int_\Omega x\mathrm{d}\mu$ 在 $\mu \in \mathcal{P}_\Omega$ 上取最大的最大值点集合是重合的. 由于 $\int_\Omega x\mathrm{d}\mu = \max_\Omega x$ 当且仅当 $\operatorname{supp}(\mu) \subset \operatorname*{argmax}_{\omega\in\Omega} x(\omega)$, 所以有

$$\partial\psi(x) = \left\{\mu \in \mathcal{P}_\Omega : \operatorname{supp}(\mu) \subset \operatorname*{argmax}_{\omega\in\Omega} x(\omega)\right\}. \tag{2.236}$$

尤其, $\partial\psi(x)$ 是单点集, 因而 ψ 在 x 处是 Hadamard 可微的, 当且仅当 $\operatorname*{argmax}_{\omega\in\Omega} x(\omega) = \{\bar{\omega}\}$ 是单点集. 此种情况 $\partial\psi(x) = \{\delta(\bar{\omega})\}$, 其中 $\delta(\omega)$ 记质量 1(mass one) 集中在点 ω 上的测度 (Dirac 测度).

凸函数 f 的次微分映射 $x \to \partial f(x)$, 具有下述重要的单调 (monotonicity) 性质. 若 $x_1, x_2 \in X$, $x_1^* \in \partial f(x_1)$, $x_2^* \in \partial f(x_2)$, 则

$$\langle x_1^* - x_2^*, x_1 - x_2\rangle \geqslant 0. \tag{2.237}$$

上述不等式通过将不等式 $\langle -x_2^*, x_1 - x_2\rangle \geqslant f(x_2) - f(x_1)$ 与 $\langle x_1^*, x_1 - x_2\rangle \geqslant f(x_1) - f(x_2)$ 相加得到, 这两个不等式由次微分 $\partial f(x_1)$ 与 $\partial f(x_2)$ 的定义得到.

例 2.123 设 S 是 X 的凸子集, 考虑相应的指示函数$I_S(x)$(见 (2.122)). 指示函数 $I_S(\cdot)$ 在点 x 处是次可微的当且仅当 $x \in S$, 此种情形, 由 (2.220) 与 (2.232) 得

$$\partial I_S(x) = \{x^* \in X^* : \sigma(x^*, S) = \langle x^*, x\rangle\} = N_S(x). \tag{2.238}$$

由 (2.237) 得, 多值函数$x \mapsto N_S(x)$ 是单调的, 这在 (2.99) 已经观察到.

现在考虑凸的正齐次函数 $\psi : X \to \overline{\mathbb{R}}$. ψ 称为是正齐次的, 若对 $\alpha > 0$, $x \in X$, 有 $\psi(\alpha x) = \alpha\psi(x)$. 当然, 对任何 $\alpha > 0$ 有 $\psi(0) = \alpha\psi(0)$, 在 $\psi(0)$ 是有限的假设下, 有 $\psi(0) = 0$. 因此, $x^* \in \partial\psi(0)$ 当且仅当

$$\psi(x) \geqslant \langle x^*, x\rangle, \quad \forall x \in X. \tag{2.239}$$

则若 $x^* \in \partial\psi(0)$, 那么 $\psi^*(x^*) = 0$, 若 $x^* \notin \partial\psi(0)$, 则 $\psi^*(x^*) = +\infty$. 即 ψ^* 等于指示函数$I_{\partial\psi(0)}$. 因此, 双共轭 $\psi^{**}(\cdot)$ 等于 $\sigma(\cdot, \partial\psi(0))$(若 $\partial\psi(0)$ 是空集, 则 $\sigma(\cdot, \partial\psi(0)) = -\infty$). 结果, 由 Fenchel-Moreau-Rockafellar 定理

$$\sigma(\cdot, \partial\psi(0)) = \text{cl}\psi(\cdot), \tag{2.240}$$

因此, $\psi(\cdot) = \sigma(\cdot, \partial\psi(0))$ 当且仅当 ψ 是闭的.

由 (2.239) 得, 若 $\partial\psi(0)$ 非空, 则 ψ 是正常的. 相反的结论在无穷维空间的情形不总是成立; 例如, 考虑例 2.114 构造的非连续的线性函数 (在有穷维空间, 有 ψ 在 0 处是次可微的当且仅当 ψ 是正常的, 见命题 2.134).

命题 2.124 令 $\psi : X \to \overline{\mathbb{R}}$是正齐次凸函数, 满足 $\psi(0) = 0$. 则下述结论成立:

(i) 下述性质是等价的:

(a) ψ 在 0 处是次可微的.

(b) ψ 在 $0 \in X$ 的邻域上是下方有界的.

(c) ψ 在 0 处是下半连续的.

(ii) 若 (a), (b) 或 (c) 中有一个成立, 则 $\sigma(\cdot, \psi(0)) = \text{lsc}\ \psi(\cdot)$.

(iii) 令 $x \in X$ 满足 $\psi(x)$ 有限. 则

$$\partial\psi(x) := \{x^* \in \partial\psi(0) : \langle x^*, x\rangle = \psi(x)\}. \tag{2.241}$$

证明 (i) 若 ψ 在 0 处是次可微的, 则存在 $x^* \in X^*$, 使 (2.239) 成立. 由于线性泛函 x^* 是连续的, 它在 0 的邻域上是有界的. 结果, 由 (a) 可得性质 (b). 由于 ψ 是正齐次的, 由 (b) 得对任意 $\varepsilon > 0$, ψ 在 0 的某一邻域上以 $-\varepsilon$ 为下界. 这表明 ψ 在 0 处是下半连续的.

现在设 ψ 在 0 处是下半连续的. 则 ψ 在 $0 \in X$ 的邻域 N 上是下方有界的. 函数 $\overline{\psi} := \text{lsc}\ \psi$ 也是正齐次的且凸的. 因为 X 是局部凸的拓扑向量空间, 邻域 N 是吸收的, 对任何 $x \in X$, 存在 x 的邻域 W 及 $\alpha > 0$ 满足 $\alpha W \subset N$. 则 $\overline{\psi}(x) > -\infty$, 因而 $\overline{\psi}$ 是正常的. 由关系式 (2.240), 有 $\overline{\psi}(\cdot) = \sigma(\cdot, \partial\overline{\psi}(0))$. 因为 $\psi(\cdot) \geqslant \overline{\psi}(\cdot)$, 得 $\partial\psi(0)$ 是非空的. 这表明 (a), (b) 与 (c) 是等价的.

(ii) 若正齐次函数 ψ 在 0 处是次可微的, 则 lsc ψ 是正常的, 因而 lsc ψ = cl ψ. 结合这一点与 (2.240), 可得到结论.

(iii) 设 $\psi(x)$ 是有限的, 令 $x^* \in \partial\psi(x)$, 即[①]

$$\psi(h) \geqslant \psi(\bar{x}) + \langle x^*, h - \bar{x}\rangle, \quad \forall h \in X. \tag{2.242}$$

因为 ψ 是正齐次的, 取 $h := t\bar{x}$, 对所有的 $t \geqslant 0$, 得到

$$0 \geqslant (1-t)[\psi(\bar{x}) - \langle x^*, \bar{x}\rangle].$$

① 从后面证明中看, 应为 $\psi(\bar{x})$ 有限, $x^* \in \partial\psi(\bar{x})$.

因为上述不等式对 $\alpha := 1 - t$ 取正值或负值均是成立的, 这表明 $\psi(\bar{x}) - \langle x^*, \bar{x}\rangle = 0$. 则由 (2.242) 得, 对所有的 $h \in X$, $\psi(h) \geqslant \langle x^*, h\rangle$, 即 $x^* \in \partial\psi(0)$. 这表明 $\partial\psi(x)$ 被包含在 (2.241) 的右端. 相反的包含关系可立即得到. □

设 $f : X \to \overline{\mathbb{R}}$ 是凸函数, 在点 $x \in X$ 处取有限值. 如在 2.2.4 节所讨论的, 由于 $t^{-1}(f(x+th) - f(x))$ 当 t 递减至 0 时是非递增的, 所以 f 在 x 处是方向可微的, 且

$$f'(x, h) = \inf_{t>0} \frac{f(x+th) - f(x)}{t}. \tag{2.243}$$

已经证得 (见命题 2.58 与 2.60)

$$\operatorname{lsc} f'(x, \cdot) = f^{\downarrow}(x, \cdot), \tag{2.244}$$

其中 $f^{\downarrow}(x, \cdot)$ 是方向上图导数(见 2.2.3 节).

下述结果可由关于凸函数方向导数的公式 (2.243) 得到.

命题 2.125 设 $f : X \to \overline{\mathbb{R}}$ 是凸函数, 在点 $x \in X$ 处取有限值, 令 $\psi(\cdot) := f'(x, \cdot)$. 则有 $\partial f(x) = \partial\psi(0)$.

证明 由 (2.243) 得

$$f(x+h) - f(x) \geqslant f'(x, h), \quad \forall h \in X,$$

因而得 $\partial\psi(0) \subset \partial f(x)$. 相反地, 若 $x^* \in \partial f(x)$, 则由定义 (2.227), 对任意 $h \in X$,

$$\frac{f(x+th) - f(x)}{t} \geqslant \langle x^*, h\rangle, \quad \forall t > 0.$$

对上述不等式取 $t \downarrow 0$ 时的极限, 得到 $\psi(h) \geqslant \langle x^*, h\rangle$, 因而 $x^* \in \partial\psi(0)$, 这证得结论. □

由上述命题得, 函数 f 在 x 处次可微当且仅当方向导数$f'(x, \cdot)$ 在 $0 \in X$ 处次可微.

命题 2.126 设 $f : X \to \overline{\mathbb{R}}$ 是凸函数, 在点 $x \in X$ 处取有限值. 则下述结果成立:

(i) f 在 x 处是次可微的当且仅当 $f'(x, \cdot)$ 在 $0 \in X$ 处是下半连续的, 即 $f^{\downarrow}(x, 0) > -\infty$, 或等价地, $f^{\downarrow}(x, 0) = 0$.

(ii) 若 f 在 x 处是次可微的, 则

$$f^{\downarrow}(x, h) = \sup_{x^* \in \partial f(x)} \langle x^*, h\rangle, \quad \forall h \in X, \tag{2.245}$$

即 $f^{\downarrow}(x, \cdot)$ 是集 $\partial f(x)$ 的支撑函数.

(iii) 若 f 在 x 处是次可微的, 则

$$N_{\mathrm{dom}\,f}(x) = [\partial f(x)]^{\infty}. \tag{2.246}$$

尤其, 若 $T_{\mathrm{dom}\,f}(x) = Y$, 则 $\partial f(x)^{\infty} = \{0\}$ ①.

(iv) 若 f 在 x 处连续, 则 f 在 x 处是次可微的.

(v) 进一步, 设 X 是赋予强拓扑的 Banach 空间. 若还有 f 在 x 处连续, 则

(a) $\partial f(x)$ 是 X^* 中的非空的凸的弱 * 紧致子集.

(b) $\partial f(x) \subset \overline{B}(0,\kappa)$, 其中 κ 是 f 在 x 点的邻域上的 Lipschitz 常数.

(c) f 在 x 处是 Hadamard 方向可微的, 且

$$f'(x,h) = \sup_{x^* \in \partial f(x)} \langle x^*, h\rangle, \quad \forall h \in X. \tag{2.247}$$

(d) 若 $\partial f(x)$ 是单点集, 即 $\partial f(x) = \{\alpha\}$, 则 f 在 x 处是 Hadamard 可微的且 $Df(x) = \alpha$.

证明　结合 (2.244), 由命题 2.124 与命题 2.125 可得命题 (i). 还有, 若 f 在 x 处是次可微的, 由命题 2.124(b), $\mathrm{lsc} f'(x,\cdot) = \sigma(\cdot,\partial f(x))$. 因为由命题 2.58 与命题 2.60, $f^{\downarrow}(x,\cdot)$ 是 $f'(x,\cdot)$ 的 l.s.c. 包, 于是 (ii) 得证.

置 $Q := \mathrm{dom}[\sigma(\cdot,\partial f(x))]$. 由命题 2.116(iii), $[\partial f(x)]^{\infty} = Q^{-}$. 因为 $\sigma(\cdot,\partial f(x)) \leqslant f'(x,\cdot)$ 且 $\mathcal{R}_{\mathrm{dom}\,f}(x) \subset \mathrm{dom}(f'(x,\cdot))$, 得 $\mathcal{R}_{\mathrm{dom}\,f}(x) \subset Q$; 因而 $[\partial f(x)]^{-} \subset \mathcal{R}_{\mathrm{dom}\,f}(x)^{-} = N_{\mathrm{dom}\,f}(x)$. 另一方面, 若 $x^* \in \partial f(x)$, $w \in N_{\mathrm{dom}\,f}(x)$, 则 $f(x') \geqslant f(x) + \langle x^* + w, x' - x\rangle$; 因而 $N_{\mathrm{dom}\,f}(x) \subset [\partial f(x)]^{\infty}$. 这就证得 (iii).

设 f 在 x 处是连续的. 则 f 在 x 的邻域 N 上是上方有界的, 即对满足 $x+h \in N$ 的所有的 h, 有 $f(x+h) \leqslant c$. 由 (2.243), 对所有的 $h \in N - x$, 有 $f'(x,h) \leqslant c - f(x)$. 即 $f'(x,\cdot)$ 在 0 的邻域上是上方有界的. 由命题 2.108, $f'(x,\cdot)$ 在 0 处是连续的, 从而由性质 (i) 得 f 在 x 处是次可微的.

现在证 (v). 设 X 是 Banach 空间, f 在 x 处是连续的. 则由命题 2.107, f 在 x 的邻域上是 Lipschitz 连续的. 则得 (见 2.2 节)$f'(x,\cdot)$ 在 X 上是 Lipschitz 连续的, 因而是下半连续的. 结果, $\partial f(x)$ 是非空的, 且由 (2.245) 得 (2.247). 进一步, 因为 f 在 x 的邻域上是以模 κ Lipschitz 连续的, 由命题 2.49 得, $f'(x,\cdot)$ 是以模 κ Lipschitz 连续的, 且 f 在 x 处是 Hadamard 方向可微的.

令 $x^* \in \partial f(x)$, 或等价地, $x^* \in \partial\psi(0)$, 其中 $\psi(\cdot) := f'(x,\cdot)$. 因为 $f'(x,\cdot)$ 是以模 κ Lipschitz 连续的, 由 (2.239) 有

$$\langle x^*, x\rangle \leqslant \kappa\|x\|, \quad \forall x \in X.$$

---

① Y 应为 X.

从而 $\|x^*\| \leqslant \kappa$, $\partial f(x) \subset \overline{B}(0,\kappa)$, 进一步, 有 $\partial f(x)$ 是 X^* 的在弱 * 拓扑下为凸的闭的. 结果, 由 Banach-Alaoglu 定理得, $\partial f(x)$ 是弱 * 紧致的. 最后, 如果 $\partial f(x)$ 是单点集合, 由 (2.247), f 是 Gâteaux 可微的. 因为在 (c) 中已经得到 Hadamard 可微性, 性质 (d) 证得. □

注 2.127 设 $f: X \to \overline{\mathbb{R}}$ 是凸函数, 在 $x \in X$ 处取有限值, $\phi(\cdot) := f^{\downarrow}(x,\cdot)$. 设 f 在 x 处不是次可微的, 因而 $\phi(\cdot)$ 在 0 处不是次可微的. 则由命题 2.126(i), $\phi(0) = -\infty$. 由 ϕ 的凸性及下半连续性得, 对所有的 $h \in T_{\operatorname{dom} f}(x)$, 有 $\phi(h) = -\infty$. 因为, 若 $h \notin T_{\operatorname{dom} f}(x)$, 容易验证 $f^{\downarrow}(x,h) = +\infty$, 于是

$$f^{\downarrow}(x,h) = \begin{cases} -\infty, & \text{若 } h \in T_{\operatorname{dom} f}(x), \\ +\infty, & \text{否则}. \end{cases} \tag{2.248}$$

例 2.128 考虑如下定义的函数 $f: \mathbb{R} \to \overline{\mathbb{R}}$,

$$f(x) := \begin{cases} +\infty, & \text{若 } x < 0, \\ -\sqrt{x}, & \text{若 } x \geqslant 0. \end{cases}$$

这一函数是凸的, l.s.c. 且是正常的. 因而得到

$$f^*(x^*) = \begin{cases} -(4x^*)^{-1}, & \text{若 } x^* < 0, \\ +\infty, & \text{若 } x^* \geqslant 0, \end{cases}$$

且

$$f'(0,h) = \begin{cases} +\infty, & \text{若 } h < 0, \\ 0, & \text{若 } h = 0, \\ -\infty, & \text{若 } h > 0. \end{cases}$$

因此, $f'(0,\cdot)$ 不是 l.s.c., 也不是正常的. 注意到

$$f^{\downarrow}(0,h) = \begin{cases} +\infty, & \text{若 } h < 0, \\ -\infty, & \text{若 } h \geqslant 0. \end{cases}$$

从而 $\operatorname{lsc} f'(0,\cdot) = f^{\downarrow}(0,\cdot)$.

例 2.129 设 X 是 Banach 空间, $f(x) = \|x\|$. 由 (2.14), $\|x\|$ 是弱 * 闭凸集 $K := \overline{B}_{X^*}$ 的支撑函数. 所以, 由命题 2.121 得

$$\partial\|x\| = \operatorname{argmax}\left\{\langle x^*, x\rangle : x^* \in \overline{B}_{X^*}\right\} = \left\{x^* \in \overline{B}_{X^*} : \langle x^*, x\rangle = \|x\|\right\}. \tag{2.249}$$

即多值映射 $x \mapsto \partial\|x\|$ 与对偶性映射是相同的, 见 2.1.3 节. 对 $x = 0$, 有 $\partial\|0\| = \overline{B}_{X^*}$, 且若 $x \neq 0$, 有

$$\partial\|x\| = \{x^* \in X^* : \|x^*\| = 1, \langle x^*, x\rangle = 1\}. \tag{2.250}$$

设对偶范数是严格凸的, 则若 $x \neq 0$, $\partial\|x\|$ 是单点集. 事实上, 若 x_1^* 与 x_2^* 是 $\partial\|x\|$ 的两个不同的元素, 则 $x^* := \frac{1}{2}(x_1^* + x_2^*)$ 属于 $\partial\|x\|$, 而 $\|x^*\| < 1$, 这是不可能的. 因此, 由命题 2.126, 若对偶范数是严格凸的, 则 $f(x) := \|x\|$ 在每一点 $x \neq 0$ 处是 Hadamard 可微的.

若 X 是 Hilbert 空间 (与其对偶相同), 对 $x \neq 0$, 由下述恒等式

$$\|x+h\|^2 = \|x^2\| + 2(x,h) + \|h\|^2,$$

相应的范数在 x 处是 Hadamard 可微的, 且 $\partial\|x\| = \{\|x\|^{-1}x\}$.

在 3.4.2 节将讨论最优化问题

$$\begin{aligned} \min \quad & f(x), \\ \text{s.t.} \quad & G(x) \in K, \end{aligned}$$

其中 K 是 Banach 空间 Y 的一闭凸子集, 引入 "不可微" 惩罚函数 $f(x) + r\,\text{dist}(G(x), K)$ 的极小化近似. 这导致下述例子的讨论.

例 2.130　设 S 是 Banach 空间X 的非空间凸子集. 对应的距离函数

$$d(x) := \text{dist}(x,\, S)$$

是模 1 Lipschitz 连续的, 当 S 是凸集时, 它是凸函数. 考虑点 $x_0 \in X$. 设存在 $\bar{x} \in S$ 满足 $d(x_0) = \|x_0 - \bar{x}\|$, 即 $\bar{x}$ 是 (可能是不唯一的) x_0 到 S 上的投影 (由推论 2.29, 若 X 是自反 Banach 空间, 则这样的 $\bar{x}$ 存在). 则得到

$$\partial d(x_0) = N_S(\bar{x}) \cap \{x^* \in X^* : x^* \in \partial\|x_0 - \bar{x}\|\}. \tag{2.251}$$

注意到, 点 $\bar{x}$ 不必是唯一的, 因此由 (2.251) 可推出, 对所有可能的 x_0 的到 S 上的投影, 这一方程 (等式) 的右端均是相同的. 注意到 $d(\cdot)$ 是凸的连续的, 则在任何情况下, 次微分$\partial d(x_0)$ 均是非空的.

下证 (2.251). 首先, 由于 $d(x)$ 是模 1 Lipschitz 连续的, 由命题 2.126(v(b)) 得 $\partial d(x_0) \subset \overline{B}_{X^*}$. 若 $x^* \in \partial d(x_0)$, 则对所有的 $x \in X$, 有

$$d(x) \geqslant d(x_0) + \langle x^*, x - x_0\rangle = \|x_0 - \bar{x}\| + \langle x^*, x - x_0\rangle. \tag{2.252}$$

在上述不等式中取 $x = \bar{x}$, 由于 $d(\bar{x}) = 0$ 及 $\|x^*\| \leqslant 1$, 得到

$$\langle x^*, x_0 - \bar{x}\rangle = \|x_0 - \bar{x}\|. \tag{2.253}$$

由于 $\partial d(x_0) \subset \overline{B}_{X^*}$, 从而得到 $x^* \in \partial\|x_0 - \bar{x}\|$(见公式 (2.249)). 结合 (2.252) 与 (2.253) 得, 对任意 $x \in X$, $d(x) \geqslant \langle x^*, x - \bar{x}\rangle$, 即 $x^* \in d(\bar{x})$. 进一步, 由于 $I_S(\bar{x}) =$

$d(\bar{x})=0$, $I_S(\cdot)\geqslant d(\cdot)$ 且 $\partial I_S(\bar{x})=N_S(\bar{x})$, 有 $\partial d(\bar{x})\subset\partial I_S(\bar{x})$, 因此 $\partial d(\bar{x})\subset N_S(\bar{x})$. 可见 $\partial d(x_0)$ 被包含在 (2.251) 的右端.

相反地, 设 x^* 属于 (2.251) 的右端. 对任何 $x\in X$ 及 $\varepsilon>0$, 存在 $x'\in S$ 满足 $d(x)\geqslant\|x-x'\|-\varepsilon$, 因此, 由 $\|x^*\|\leqslant 1$, $x^*\in N_S(\bar{x})$, 可得

$$d(x)\geqslant\langle x^*,x-x'\rangle-\varepsilon=\langle x^*,x-\bar{x}\rangle+\langle x^*,\bar{x}-x'\rangle-\varepsilon\geqslant\langle x^*,x-\bar{x}\rangle-\varepsilon.$$

因为 ε 是任意的, 由 (2.249) 可得

$$d(x)\geqslant\langle x^*,x-\bar{x}\rangle=\langle x^*,x-x_0\rangle+\langle x^*,x_0-\bar{x}\rangle=d(x_0)+\langle x^*,x-x_0\rangle,$$

从而得 $x^*\in\partial d(x_0)$①. 这就证得 (2.251).

尤其, 若 $x_0\in S$, 则 $\bar{x}=x_0$, 且由于 $\partial\|0\|=\overline{B}_{X^*}$, 公式 (2.251) 具有下述形式

$$\partial d(x_0)=N_S(x_0)\cap\overline{B}_{X^*},\quad 若\ x_0\in S. \tag{2.254}$$

再注意到, 若 $x_0\notin S$, 则 $d(x_0)>0$, 对偶范数是严格凸的, 因而 $\partial d(x_0)$ 是单点集, $d(x)$ 在 x_0 处是 Hadamard 可微的. 尤其, 若 X 是 Hilbert 空间, 则由于到 S 上的投影 P_S 是单值映射, 且 $\partial\|x\|=\{\|x\|^{-1}x\}$, 有 $d(\cdot)$ 在每一 $x_0\notin S$ 处是 Hadamard 可微的, 其导数为 $(x_0-P_S(x_0))/\|x_0-P_S(x_0)\|$.

公式 (2.246) 容许我们给出 $\partial f(x)$ 的有界性的如下刻画. 若 Banach 空间X 的凸子集是有界的, 则其回收锥是 $\{0\}$. 而这一结论的相反命题一般来说是不真的 (例 2.43), 除非空间 X 是有限维的.

命题 2.131 设 X 是局部凸的拓扑向量空间, $f:X\to\overline{\mathbb{R}}$ 是凸函数, 它在点 $x\in X$ 处取有限值. 若 $\partial f(x)$ 非空, $\partial f(x)$ 的回收锥是 $\{0\}$ 且 $\operatorname{dom} f$ 具有非空的相对内部, 则 $x\in\operatorname{int}(\operatorname{dom} f)$. 相反地, 若 X 是 Banach 空间, $x\in\operatorname{int}(\operatorname{dom} f)$, 则 $\partial f(x)$ 是有界的.

证明 因为 $\partial f(x)$ 非空, 回收锥是 $\{0\}$, 由 (2.246) 得 $N_{\operatorname{dom} f}(x)=\{0\}$, 因而 $T_{\operatorname{dom} f}(x)=X$. 这表明 $\operatorname{Sp}(\operatorname{dom} f-x)$ 在 X 中是稠密的, x 不能从 $\operatorname{dom} f$ 分离. 由假设 $\operatorname{dom} f$ 具有非空的相对内部得 $\operatorname{dom} f$ 的内部是非空的, 由第一分离定理 2.13 得 $x\in\operatorname{int}(\operatorname{dom} f)$ 当且仅当 x 不能从 $\operatorname{dom} f$ 分离开. 由 $T_{\operatorname{dom} f}(x)=X$ 得 $x\in\operatorname{int}(\operatorname{dom} f)$.

相反地, 设 X 是 Banach 空间, $x\in\operatorname{int}(\operatorname{dom} f)$. 若 $\partial f(x)$ 是空集, 当然, $\partial f(x)$ 是有界的. 因此设 f 在 x 处是次可微的. 考虑 $\psi(\cdot):=f^{\downarrow}(x,\cdot)$. 因为 $x\in\operatorname{int}(\operatorname{dom} f)$, 则对所的 $h\in X$, $f^{\downarrow}(x,h)<+\infty$, 即 $0\in\operatorname{int}(\operatorname{dom}\psi)$. 由于 ψ 是下半连续的, f 是 x 点处次可微的, ψ 是正常的. 则由命题 2.111 得 ψ 在 0 处是连续的. 结果, 由命题 2.126(v), $\partial\psi(0)$ 亦即 $\partial f(x)$ 是有界的. □

① 原著中为 $x^*\in d(x_0)$.

若空间 X 是有限维的, 则 X 的每一凸子集具有非空的相对内部, 定义在 X 上的凸函数在点 $x \in \operatorname{dom} f$ 处是连续的, 满足 $f(x) > -\infty$ 当且仅当 $x \in \operatorname{int}(\operatorname{dom} f)$. 所以, 在有限维情形, 下述结论成立.

命题 2.132 设 X 是有限维空间的, $f: X \to \overline{\mathbb{R}}$ 是凸函数, 在 $x \in X$ 处取有限值. 则 $\partial f(x)$ 非空且有界的充分必要条件是 f 在 x 处连续.

下述例子表明, 在无穷维空间的情况下, 命题 2.131 中 $\operatorname{dom} f$ 相对内部非空性这一条件是本质性的.

例 2.133 考虑空间 $X := L_2[0,1]$, $K \subset L_2[0,1]$ 是几乎处处非负值函数的集合, 指示函数 $f(\cdot) := I_K(\cdot)$. 由于 K 是 $L_2[0,1]$ 的非空闭凸子集, 指示函数 f 是凸的且是下半连续的. 考虑 $x(t) = 1$, $\forall t \in [0,1]$. 得 $x \in L_2[0,1]$, $T_K(x) = L_2[0,1]$, 且 $\partial f(x) = \{0\}$(如例 2.64). 这说明, $\partial f(x)$ 是非空的且有界的. 然而, x 不是 K 的内部点 (K 的内部是空集), 且 f 在 x 处是非连续的.

命题 2.134 设 $f: X \to \overline{\mathbb{R}}$ 是凸函数, 在 x_0 处取有限值. 则下述成立:

(i) 若 f 在 x_0 处是次可微的, 则对任意 $h \in X$, $f'(x_0, h) > -\infty$.

(ii) 相反地, 若在一点 $\bar{x} \in X$ 的邻域上, $f(\cdot)$ 是上方有界的, 若对所有的 $h \in X$, $f'(x_0, h) > -\infty$, 则 $f(\cdot)$ 于 x_0 处是次可微的.

(iii) 若空间 X 是有限维的, 则 $f(\cdot)$ 在 x_0 处是次可微的当且仅当对所有的 $h \in X$ 有 $f'(x_0, h) > -\infty$.

(iv) 若空间 X 是有限维的且 $x \in \operatorname{ri}(\operatorname{dom} f)$, 则 f 在 x 处是次可微的.

证明 若 $x^* \in \partial f(x_0)$, 则对任何 $h \in X$, $f'(x_0, h) \geqslant \langle x^*, h \rangle > -\infty$. 这证得 (i). 相反地, 设 $f(\cdot)$ 在 $\bar{x}$ 的邻域上是上方有界的, 且对所有的 $h \in X$, $f'(x_0, h) > -\infty$. 若 $x_0 = \bar{x}$, 则由命题 2.107 与 2.126(v), $f(\cdot)$ 在 x_0 处是连续的, 因而在 x_0 处是次可微的. 设 $x_0 \neq \bar{x}$. 因 $f(\cdot) - f(x_0)$ 在 $\bar{x}$ 点的邻域上是上方有界的, 由 (2.243), $\psi(\cdot) := f'(x_0, \cdot)$ 在点 $\bar{x} - x_0$ 的某一邻域上是上方有界的. 这同 $\psi(\cdot) > -\infty$ 可推出 $\bar{x} \in \operatorname{dom} \psi$. 因 $\psi(\cdot)$ 是凸的, 由命题 2.107 与 2.126(v) 得 $\psi(\cdot)$ 是连续, 因此在 $\bar{x}$ 处有次梯度x^*. 由命题 2.124(c) 得 $x^* \in \partial f'(x_0, \cdot)$, 结论由命题 2.125 得到.

现在设空间 X 是有限维的. 此种情形, 集合 $\operatorname{dom} f$ 具有非空的相对内部, 限定 f 在由 $\operatorname{dom} f$ 生成的仿射空间上, 如有必要, 不妨设 $\operatorname{dom} f$ 具有非空的内部. 因为 X 是有限维的, 则 f 在 $\operatorname{dom} f$ 之内部的任何点均是连续的. 则结论 (iii) 由 (i) 与 (ii) 得到.

注意到 (iii), 为证 (iv), 只需证若 $x \in \operatorname{ri}(\operatorname{dom} f)$, 则 $f'(x, \cdot)$ 是正常的. 设 h 是给定的方向. 若 $h \notin \mathcal{R}_{\operatorname{dom} f}(x)$, 则对任何 $t > 0$, $f(x + th) = +\infty$, 因而 $f'(x, h) = +\infty$. 设 $h \in \mathcal{R}_{\operatorname{dom} f}(x)$, 考虑函数 $\phi(t) := f(x + th)$, 它是凸函数. 因为 $x \in \operatorname{ri}(\operatorname{dom} f)$, 有 $-h \in \mathcal{R}_{\operatorname{dom} f}(x)$. 结果, 存在正数 t_1, t_2 满足值 $f(x + t_1 h)$ 与 $f(x - t_2 h)$ 是有限的, 因而 ϕ 在包含 0 的区间上有有限值. 因此 ϕ 在零点处是连续的, 且 $\phi'(0, 1)$ 是有限

的. 由 $\phi'(0,1) = f'(x,h)$, 这完成证明. □

应该指出, 在无穷维空间的情形, 仅有条件 $f'(x_0,h) > -\infty, \forall h \in X$ 不能推出 f 在 x_0 处的次可微性. 例如, 考虑非连续的线性函数 $f: X \to \mathbb{R}$. 所以, 在命题 2.134(ii) 中, f 在某邻域上是上方有界这一假设是本质性的.

2.4.4 链式法则

这一节, 我们讨论复合函数可微性的链式法则. 设 X 与 Y 是 Banach 空间. 关于切锥的链式法则容易由稳定性定理 2.87 (见推论 2.91) 得到.

令 $G: X \to Y$ 是连续可微的映射, $K \subset Y$ 是非空闭凸集, 考虑 $\Phi := G^{-1}(K)$ 及点 $x_0 \in \Phi$. 设 Robinson 约束规范 (2.178) 成立. 则 Φ 在 x_0 点的余切锥与内切锥是重合的, 且

$$T_\Phi(x_0) = (DG(x_0))^{-1}[T_K(G(x_0))]. \tag{2.255}$$

注 2.135 上述命题的结果可表示为下述可交换的图：

$$\begin{array}{ccc} \Phi & \xrightarrow{G} & K \\ \big\downarrow & & \big\downarrow \\ T_\Phi & \xrightarrow{DG} & T_K \end{array}$$

其中竖直箭头表示切集运算.

上述结果可解释为相应的复合函数的链式法则.

命题 2.136 设 $G: X \to Y$ 是连续可微映射, $f: Y \to \overline{\mathbb{R}}$ 是 l.s.c. 凸函数, 在点 $y_0 := G(x_0)$ 处取有限值. 设关于集合 $K := \operatorname{dom} f$ 于 x_0 处 Robinson 约束规范(2.178) 成立. 则复合函数 $f \circ G$ 在 x_0 处的下, 上方向上图导数是相同的, 且

$$(f \circ G)^{\downarrow}(x_0,h) = f^{\downarrow}(G(x_0), DG(x_0)h). \tag{2.256}$$

证明 令 $\widehat{K} := \operatorname{epi} f$, $\widehat{G}(x,\alpha) := (G(x),\alpha)$, $\alpha \in \mathbb{R}$. 不难验证 $\widehat{G}^{-1}(\widehat{K}) = \operatorname{epi}(f \circ G)$, 由 (2.178) 可推出集合 $\widehat{K}$ 与 $\widehat{G}$ 在 $(x_0, f(y_0))$ 处的 Robinson 约束规范成立, 即

$$0 \in \operatorname{int}\left\{ \begin{bmatrix} G(x_0) \\ f(G(x_0)) \end{bmatrix} + \begin{bmatrix} DG(x_0)X \\ \mathbb{R} \end{bmatrix} - \operatorname{epi}(f) \right\}.$$

由方程 (2.255) 得到

$$T_{\operatorname{epi}(f\circ G)}[x_0, (f \circ G)(x_0)] = D\widehat{G}(x_0,\alpha)^{-1} T_{\operatorname{epi}(f)}[\widehat{G}(x_0, f(y_0))]. \tag{2.257}$$

结合命题 2.58 即得结论. □

为将上述结果应用到与凸函数的和相联系的复合函数上, 需要下述引理.

引理 2.137　设 $f: X \to \overline{\mathbb{R}}$ 与 $g: Y \to \overline{\mathbb{R}}$ 是正常凸函数, 考虑复合函数 $F(x,y) := f(x) + g(y)$, 其定义域是

$$\operatorname{dom} F = \operatorname{dom} f \times \operatorname{dom} g.$$

则 F 在 $(x,y) \in \operatorname{dom} F$ 处是方向上图可微的, 且

$$F^{\downarrow}((x,y),(h,d)) = \begin{cases} +\infty, & (h,d) \notin T_{\operatorname{dom} f}(x) \times T_{\operatorname{dom} g}(y), \\ f^{\downarrow}(x,h) + g^{\downarrow}(y,d), & (h,d) \in T_{\operatorname{dom} f}(x) \times T_{\operatorname{dom} g}(y) \\ & \text{且 } \partial f(x) \neq \varnothing, \partial g(y) \neq \varnothing, \\ -\infty, & \text{否则}. \end{cases}$$

证明　由于 $F^{\downarrow}((x,y),(\cdot,\cdot)) = \operatorname{lsc} F'((x,y),(\cdot,\cdot))$, 又

$$F^{\downarrow}((x,y),(h,d)) = \begin{cases} f'(x,h) + g'(y,d), & (h,d) \in \mathcal{R}_{\operatorname{dom} f}(x) \times \mathcal{R}_{\operatorname{dom} g}(y), \\ +\infty, & \text{否则}, \end{cases}$$

显然, 若 $(h,d) \notin T_{\operatorname{dom} f}(x) \times T_{\operatorname{dom} g}(y)$, 有 $F^{\downarrow}((x,y),(h,d)) = +\infty$.

现在设 $(h,d) \in T_{\operatorname{dom} f}(x) \times T_{\operatorname{dom} g}(y)$. 考虑 $\partial f(x)$ 与 $\partial g(y)$ 非空集的情况. 因为 $f'(x,h) + g'(y,d)$ 是下方有界的, 由上述等式得 $F^{\downarrow}((x,y),(h,d)) = f^{\downarrow}(x,h) + g^{\downarrow}(y,d) > -\infty$. 若 $\partial f(x)$ 是空集, 由注 2.127, 对所有的 $h \in T_{\operatorname{dom} f}(x)$, $f^{\downarrow}(x,h) = -\infty$. 对 $g(\cdot)$ 也有类似的结论, 则若 $\partial f(x)$ 或者 $\partial g(y)$ 是空集, 有 $F^{\downarrow}((x,y),(\cdot,\cdot))$ 在 $T_{\operatorname{dom} f}(x) \times T_{\operatorname{dom} g}(y)$ 上取值 $-\infty$, 否则取 $+\infty$. □

结合前面的两个结果可得到下述结论.

定理 2.138　设 $f: X \to \overline{\mathbb{R}}$ 与 $g: Y \to \overline{\mathbb{R}}$ 是正常的 l.s.c. 凸函数, 令 $A \in \mathcal{L}(X,Y)$. 考虑复合函数 $F(x) := f(x + g(Ax))$, 其中

$$\operatorname{dom}(F) = \{x \in \operatorname{dom} f : Ax \in \operatorname{dom}(g)\}.$$

设下述正则性条件

$$0 \in \operatorname{int}\{A \operatorname{dom} f - \operatorname{dom} g\} \tag{2.258}$$

成立. 则对任何 $x_0 \in \operatorname{dom} F$,

$$F^{\downarrow}(x_0,h) = \begin{cases} +\infty, & \text{若 } (h,Ah) \notin T_{\operatorname{dom} f}(x_0) \times T_{\operatorname{dom} g}(Ax_0), \\ f^{\downarrow}(x_0,h) + g^{\downarrow}(Ax_0,Ah), & \text{若 } h \in T_{\operatorname{dom} f}(x_0) \cap A^{-1}[T_{\operatorname{dom} g}(Ax_0)] \\ & \text{且 } \partial f(x_0) \neq \varnothing, \partial g(Ax_0) \neq \varnothing, \\ -\infty, & \text{否则}. \end{cases}$$

证明 设 $F(x)$ 是映射 $G: x \to (x, Ax)$ 与函数 $(x,y) \mapsto f(x)+g(y)$ 的复合. 约束规范 (2.202) 取下述形式 (比较 (2.194))

$$0 \in \operatorname{int}\Big\{G(X) - \operatorname{dom}(f) \times \operatorname{dom}(g)\Big\},$$

即它等价于 (2.258). 根据引理 2.137 相对应的公式, 由命题 2.136 的链式法则即得结果. □

注 2.139 在定理 2.138 的假设下, 若还有 F 在 x_0 处是次可微的, 则 $F^{\downarrow}(x_0, 0) > -\infty$, 因而由命题 2.126(i) 得 $f^{\downarrow}(x_0, 0) > -\infty$, $g^{\downarrow}(Ax_0, 0) > -\infty$. 由命题 2.126(i), f (与 g) 在 x_0 (与 $A(x_0)$) 是次可微的. 结合命题 2.126(ii), 对所有的 $h \in X$, 得①

$$\begin{aligned}\sup_{x^* \in \partial F(x)} \langle x^*, h\rangle &= \sup_{x^* \in \partial f(x)} \langle x^*, h\rangle + \sup_{y^* \in \partial g(Ax)} \langle y^*, Ah\rangle \\ &= \sup\{\langle x^*, h\rangle : x^* \in \partial f(x) + A^* \partial g(Ax)\}.\end{aligned}$$

由于 $\partial F(x_0)$ 是弱 * 闭凸集, 由命题 2.116 得

$$\partial F(x_0) = \mathrm{cl}^*\{\partial f(x_0) + A^* \partial g(Ax_0)\}. \tag{2.259}$$

2.5.4 节将要证明等式 $\partial F(x_0) = \partial f(x_0) + A^* \partial g(Ax_0)$ 在条件 (2.258) 下成立.

例 2.140 该例说明 $f^{\downarrow}(x_0, h)$ 或 $g^{\downarrow}(Ax_0, Ah)$ 取无穷值的情况. 令 $X = Y = \mathbb{R}^2$,

$$f(x) := \begin{cases} +\infty, & \text{若} x_1 < 0, \\ -\sqrt{x_1}, & \text{若} x_1 \geqslant 0, \end{cases}$$

$$g(x) := \begin{cases} 0, & \text{若 } x_1 = 0, \\ (x_1)^2/x_2, & \text{若 } x_1 \geqslant 0, x_2 > 0, \\ +\infty, & \text{否则}. \end{cases}$$

这是两个 l.s.c. 正常的凸函数. 令 A 是单位映射, 即 $Ax = x$, $\forall x \in \mathbb{R}^2$. 由于 $\operatorname{dom} f - \operatorname{dom} g = \mathbb{R}^2$, 正则条件成立. 由推论 2.138 得, $F(x) := f(x) + g(x)$ 满足对下方向上图导数, 链式法则 (2.256) 成立. 置 $x_0 = 0$. 则

$$F^{\downarrow}(0, h) := \begin{cases} +\infty, & \text{若 } h_1 < 0 \text{ 或 } h_2 < 0, \\ -\infty, & \text{否则}. \end{cases}$$

当 $h_1 > 0$ 且 $h_2 < 0$, $f^{\downarrow}(0, h) = -\infty$ 且 $g^{\downarrow}(0, h) = +\infty$, 而 $F^{\downarrow}(0, h) = +\infty$. 对 $h = (1, 0)$, $f^{\downarrow}(0, h) = -\infty$ 且 $g^{\downarrow}(0, h) = 0$, 而 $F^{\downarrow}(0, h) = -\infty$.

① 下面的 x 应为 x_0.

2.5 对偶理论

2.5.1 共轭对偶性

这一节讨论最优化问题的对偶理论中的共轭 (或参数) 方式. 除非特别说明, 设 (X, X^*), (U, U^*) 与 (Y, Y^*) 均是成对的空间 (见定义 2.26), 即每一对中的一个空间是局部凸的拓扑向量空间, 而另一空间是这一空间的拓扑对偶, 在 X, U 或 Y 是 Banach 空间的情形, 赋予它强拓扑, 而其对偶空间赋予成对的拓扑 (paired topology). 例如, 若 X 是自反的 Banach 空间, 赋 X^* 以强 (对偶) 拓扑, 否则赋予弱 * 拓扑.

考虑最优化问题

$$\text{(P)} \qquad \min_{x \in X} f(x), \tag{2.260}$$

其中 $f : X \to \overline{\mathbb{R}}$. 设问题 (P) 被嵌入到下述参数类优化问题

$$(\mathrm{P}_u) \qquad \min_{x \in X} \varphi(x, u), \tag{2.261}$$

其中 $\varphi : X \times U \to \overline{\mathbb{R}}$, 设 $u = 0$ 对应的问题 (P_0) 与 (P) 相同, 即 $\varphi(\cdot, 0) = f(\cdot)$, 此时不限定 f 或 φ 是凸函数.

与原始问题 (P_u) 相联系的最优值函数是

$$v(u) := \inf_{x \in X} \varphi(x, u). \tag{2.262}$$

称问题 (P)(问题 (P_u)) 是相容的, 若 $f(\cdot)$ 的定义域 ($\varphi(\cdot, u)$ 的定义域) 是非空的. 显然, (P_u) 是相容的当且仅当 $v(u) < +\infty$.

基于下述共轭函数来计算 $v(\cdot)$ 的共轭函数

$$\varphi^*(x^*, u^*) = \sup_{(x,u) \in X \times U} \{\langle x^*, x\rangle + \langle u^*, u\rangle - \varphi(x, u)\}, \tag{2.263}$$

有

$$\begin{aligned} v^*(u^*) &= \sup_{u \in U} \{\langle u^*, u\rangle - v(u)\} \\ &= \sup_{u \in U} \{\langle u^*, u\rangle - \inf_{x \in X} \varphi(x, u)\} \\ &= \sup_{(x,u) \in X \times U} \{\langle u^*, u\rangle - \varphi(x, u)\} = \varphi^*(0, u^*). \end{aligned} \tag{2.264}$$

显然有

$$v^{**}(u) = \sup_{u^* \in U^*} \{\langle u^*, u\rangle - \varphi^*(0, u^*)\}. \tag{2.265}$$

这导致下述定义的对偶问题(dual problem)(即原始问题 (P_u) 的对偶)

$$(\mathrm{D}_u) \qquad \max_{u^*\in U^*}\{\langle u^*,u\rangle-\varphi^*(0,u^*)\}. \tag{2.266}$$

称上述问题 (D_u) 是 (P_u) 的共轭对偶(conjugate dual). 尤其, 对 $u=0$, 对应的问题 (D_0) 为

$$(\mathrm{D}) \qquad \max_{u^*\in U^*}\{-\varphi^*(0,u^*)\}, \tag{2.267}$$

它被视为 (P) 的 (共轭对偶). 显然有 $\mathrm{val}(\mathrm{P}_u)=v(u)$, $\mathrm{val}(\mathrm{D}_u)=v^{**}(u)$, 因为 $v(u)\geqslant v^{**}(u)$, 有

$$\mathrm{val}(\mathrm{P}_u)\geqslant \mathrm{val}(\mathrm{D}_u). \tag{2.268}$$

非负量 $\mathrm{val}(\mathrm{P}_u)-\mathrm{val}(\mathrm{D}_u)$, 当它有意义的时候, 被称为是 (P_u) 与 (D_u) 间的对偶间隙(duality gap).

命题 2.141 *若 $v^{**}(u)$ 是有限的, 则 (D_u) 的最优解集 $\mathcal{S}(\mathrm{D}_u)$ 与 $\partial v^{**}(u)$ 相同.*

证明 由 (2.265) 与命题 2.118 得到. □

定理 2.142 *下述结论成立:*

(i) *若对给定的 $u\in U$, 次微分$\partial v(u)$ 是非空的, 则 (P_u) 与 (D_u) 的对偶间隙为零, 即 $\mathrm{val}(\mathrm{P}_u)=\mathrm{val}(\mathrm{D}_u)$, 对偶问题$(\mathrm{D}_u)$ 的最优解集与 $\partial v(u)$ 相同.*

(ii) *若 $\mathrm{val}(\mathrm{P}_u)=\mathrm{val}(\mathrm{D}_u)$ 且有限, 则 (D_u) 的最优解集 (可能是空集) 与 $\partial v(u)$ 相同.*

(iii) *若 $\mathrm{val}(\mathrm{P}_u)=\mathrm{val}(\mathrm{D}_u)$ 且 $\bar{x}\in X$ 与 $\bar{u}^*\in U^*$ 分别是 (P_u) 与 (D_u) 的最优解 (从而公共最优值是有限的), 则下述最优性条件*

$$\varphi(\bar{x},u)+\varphi^*(0,\bar{u}^*)=\langle\bar{u}^*,u\rangle \tag{2.269}$$

成立. 相反地, 若条件 (2.269) 对某 $\bar{x}$ 与 $\bar{u}^*$ 成立, 则 $\bar{x}$ 与 $\bar{u}^*$ 分别是 (P_u) 与 (D_u) 的最优解, (P_u) 与 (D_u) 不存在对偶间隙.

证明 由 Young-Fenchel 不等式, 若 $\partial v(u)$ 是非空的, 则

$$v(u)\geqslant\langle u^*,u\rangle-v^*(u^*),$$

且由 (2.229),

$$v(u)=\langle u^*,u\rangle-v^*(u^*)\ \text{当且仅当}\ u^*\in\partial v(u).$$

因为 $v^*(u^*)=\varphi^*(0,u^*)$, 这意味着 u^* 是 (D_u) 的最优解的充分必要条件是 $u^*\in\partial v(u)$, 此种情况 $\mathrm{val}(\mathrm{P}_u)=\mathrm{val}(\mathrm{D}_u)$, 相反的结论也是成立的. 因此得到 (i) 与 (ii). 同样还可由命题 2.141, $\mathcal{S}(\mathrm{D}_u)=\partial v^{**}(u)$. 因为 $v(\cdot)$ 在 u 处是次可微的, 由命题 2.118

得, $\partial v^{**}(u) = \partial v(u)$, 从而得到 (i). 若 $v^{**}(u) = v(u)$, 则 $\partial v^{**}(u) = \partial v(u)$, 从而得到 (ii).

若 $\mathrm{val}(\mathrm{P}_u) = \mathrm{val}(\mathrm{D}_u)$ 是有限的, 由上述讨论, 显然 $\bar{x} \in X$, $\bar{u}^* \in U^*$ 分别是 (P_u) 与 (D_u) 的最优解当且仅当条件 (2.269) 成立. 显然, 若 (2.269) 对 $\bar{x}$ 与 $\bar{u}^*$ 成立, 则 $\mathrm{val}(\mathrm{P}_u) = \mathrm{val}(\mathrm{D}_u)$, (iii) 得证. □

由定理 2.142(ii), 若存在 $u \in U$, (P_u) 与 (D_u) 间没有对偶间隙, 则最优值函数$v(\cdot)$ 在 u 处是次可微的当且仅当对偶问题 (D_u) 的最优解集 $\mathcal{S}(\mathrm{D}_u)$ 是非空的. 进一步, 若 $v(\cdot)$ 是凸的, 在 u 处是次可微的, 则它的方向导数满足

$$v'(u,d) \geqslant (v^{**})'(u,d) \geqslant \sup_{u^* \in \mathcal{S}(\mathrm{D}_u)} \langle u^*, d\rangle, \quad \forall d \in U, \tag{2.270}$$

再由命题 2.126, 若还有 $v'(u,\cdot)$ 在 0 处是下半连续的, 则上面公式中等式成立.

下述结果表明, 若函数 $\varphi(x,u)$ 是凸的 (作为定义在空间 $X \times U$ 上的增广实值函数), 则最优值函数 $v(u)$ 是凸的.

命题 2.143 若函数 $\varphi(x,u)$ 是凸的, 则最优值函数 $v(u)$ 是凸的.

证明 若 $\varphi(x,u)$ 是凸的, 则对任意 $x_1, x_2 \in X, u_1, u_2 \in U, t \in [0,1]$,

$$\begin{aligned} t\varphi(x_1,u_1) + (1-t)\varphi(x_2,u_2) &\geqslant \varphi(tx_1 + (1-t)x_2, tu_1 + (1-t)u_2) \\ &\geqslant v(tu_1 + (1-t)u_2), \end{aligned}$$

在左端对 x_1 与 x_2 极小化, 得

$$tv(u_1) + (1-t)v(u_2) \geqslant v(tu_1 + (1-t)u_2),$$

从而证得 $v(u)$ 是凸的. □

注意到, 即使 $\varphi(x,u)$ 是凸的正常的下半连续函数, 也可能发生最优值函数 $v(u)$ 非正常的或非下半连续的情况.

现在设 φ 是凸的, 因此 v 也是凸的. 由 Fenchel-Moreau-Rockafellar 定理 2.113, $v^{**} = \mathrm{cl}v$,

$$\mathrm{cl}v(\cdot) := \begin{cases} \mathrm{lsc}\ v(\cdot), & \text{若 } \mathrm{lsc}\ v(u) > -\infty, \forall u \in U, \\ -\infty, & \text{若至少存在 } u \in U, \text{ 满足 } \mathrm{lsc}\ v(u) = -\infty, \end{cases} \tag{2.271}$$

且

$$\mathrm{lsc}\ v(u) = \min\{v(u),\ \liminf_{u' \to u} v(u')\}. \tag{2.272}$$

进一步, 若 v 的定义域中含有至少两个不同的点, 则

$$\mathrm{lsc}\ v(u) = \liminf_{u' \to u} v(u'). \tag{2.273}$$

值 lsc $v(u)$ 称为问题 (P_u) 的次值(subvalue). 问题 (P_u) 称为是次相容的 (subconsistent). 若次值小于 $+\infty$. 由 (2.272) 得, (P_u) 是次相容的当且仅当在 u 的任何邻域内存在一点 u', 满足 $\mathrm{val}(P_{u'})$ 是上方有界的. 若 v 是凸函数, 则由命题 2.110, lsc v 是正常的, 若它在 U 中的至少一点处取有限值. 否则, lsc v 只取两个值: $+\infty$ 或 $-\infty$. 因此, 只有下述三种可能性之一发生: (i) lsc v 是正常的, 因此 lsc $v(\cdot) \equiv \mathrm{cl}\, v(\cdot)$; (ii) lsc $v(\cdot) \equiv \mathrm{cl}\, v(\cdot) = +\infty$; (iii) $\mathrm{cl}\, v(\cdot) \equiv -\infty$. 若 (P_u) 是次相容的, 则任意情况均有 lsc $v(u) = \mathrm{cl}\, v(u)$. 得到下述结果.

定理 2.144 若函数 $\varphi(x,u)$ 是凸的, 则

$$\mathrm{val}(\mathrm{D}_u) = \mathrm{cl}\, v(u). \tag{2.274}$$

进一步, 若还有问题 (P_u) 是次相容的, 则

$$\mathrm{val}(\mathrm{D}_u) = \mathrm{lsc}\, v(u) = \min\left\{v(u), \liminf_{u' \to u} v(u')\right\}. \tag{2.275}$$

由定理 2.144 得, 若 φ 是凸的, (P_u) 是次相容的, 则 (P_u) 与 (D_u) 间没有对偶间隙当且仅当最优值函数 $v(\cdot)$ 在点 u 处是下半连续的, 即

$$v(u) \leqslant \liminf_{u' \to u} v(u'). \tag{2.276}$$

例 2.145 令

$$\varphi(x,u) := \begin{cases} f(x), & \text{若 } \mathrm{e}^x - u \leqslant 0, \\ +\infty, & \text{否则}, \end{cases}$$

其中 $x, u \in \mathbb{R}$ 且 $f(x)$ 是实值凸函数. 则 $\varphi : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ 是凸的正常的下半连续的.

(i) 令 $f(x) := x$. 则

$$v(u) = \begin{cases} +\infty, & \text{若 } u \leqslant 0, \\ -\infty, & \text{若 } u > 0, \end{cases}$$

且 $v^{**}(u) \equiv -\infty$. 此种情形, $v(u)$ 不是正常的, 亦不是闭的. 进一步, $v(0) = +\infty$(问题 (P) 不是相容的), $v^{**}(0) = -\infty$, 因此问题 (P) 与 (D) 间的对偶间隙是无穷大. 另一方面, 问题 (P) 是次相容的, 它的次值是 $-\infty$. 对 $u < 0$, 问题 (P_u) 不是次相容的, 其次值是 $+\infty$, 对 (D_u) 其最优值是 $-\infty$.

(ii) 令 $f(x) := \mathrm{e}^x$. 则

$$v(u) = \begin{cases} +\infty, & \text{若 } u \leqslant 0, \\ 0, & \text{若 } u > 0. \end{cases}$$

此种情况, $v(u)$ 是正常的, 但不是下半连续的. 也有

$$\varphi^*(0,u^*) = v^*(u^*) = \begin{cases} 0, & \text{若 } u^* \leqslant 0, \\ +\infty, & \text{若 } u^* > 0, \end{cases}$$

$$v^{**}(u) = \begin{cases} +\infty, & \text{若 } u < 0, \\ 0, & \text{若 } u \geqslant 0. \end{cases}$$

可见, 对 $u = 0, \text{val(P)} = v(0) = +\infty$, 而val(D)=0. 即 (P) 与 (D) 间的对偶间隙是 $+\infty$. 问题 (P) 是次相容的, 其次值是 0.

定义 2.146 称问题 (P_u) 是平静的 (calm), 若 $\text{val}(\text{P}_u)$ 有限且最优值函数$v(\cdot)$ 在 u 处是次可微的, 即 $\partial v(u) \neq \varnothing$.

由定理 2.142, 得到下述结论.

命题 2.147 设 $\text{val}(\text{P}_u)$ 是有限的. 若 (P_u) 是平静的, 则 (P_u) 与 (D_u) 间不存在对偶间隙, 对偶问题 (D_u) 的最优解集非空. 相反地, 若 (P_u) 与 (D_u) 间不存在对偶间隙, 则对偶问题(D_u) 有最优解的充分必要条件是 (P_u) 是平静的.

在某些特殊情况, 验证最优值函数 $v(\cdot)$ 的次可微性, 因而原问题 (P_u) 的平静性是可能的. 下述的结果与第 3 章中精确罚函数的讨论相关.

命题 2.148 设最优值函数 $v(\cdot)$ 是凸的, $\text{val}(\text{P}_u)$ 是有限的. 则下述结论成立:

(i) 设 X 是 Banach 空间. 则 (P_u) 是平静的当且仅当存在正数 r 满足函数 $\theta_r(\cdot) := v(\cdot) + r\|\cdot - u\|$ 在 u 处达到最小.

(ii) 若最优值函数 $v(\cdot)$ 在点 $\bar{u} \in U$ 的邻域上是上方有界的或空间 U 是有限维的, 则 (P_u) 是平静的充要条件是下述条件成立

$$\liminf_{t \downarrow 0} \frac{v(u+td) - v(u)}{t} > -\infty, \quad \forall d \in U. \tag{2.277}$$

(iii) 若空间 U 是有限维的, $u \in \text{ri}\ (\text{dom} v)$, 则 (P_u) 是平静的.

证明 设 X 是 Banach 空间, 函数 $\theta_r(\cdot)$ 在 u 处取极小. 由函数 $\theta_r(\cdot)$ 是凸函数, 得 $\theta_r'(u,\cdot) \geqslant 0$. 因为 $\|\cdot\|$ 是正齐次的连续的, 得 $v'(u,\cdot)$ 在 0 点的邻域上是上方有界的, 因而于 0 处是 l.s.c.的. 由命题 2.126 得, $v(\cdot)$ 在 u 处是次可微的, 因而 (P_u) 是平静的. 相反地, 设 $v(\cdot)$ 在 u 处是次可微的, $u^* \in \partial v(u)$. 对任何 $u' \in U$, 得

$$\theta_r(u') \geqslant \theta_r(u) + \langle u^*, u' - u\rangle + r\|u' - u\| \geqslant \theta_r(u) + (r - \|u^*\|)\|u' - u\|,$$

因而对任意的 $r \geqslant \|u^*\|$, $\theta_r(\cdot)$ 在 u 处取最小值. 这证得 (i).

结论 (ii) 与 (iii) 由命题 2.134 的相应的结果得到. □

注 2.149 注意到不需要 $v(\cdot)$ 的凸性, 有结论: "若 $v(\cdot)$ 在 u 处是次可微的, 则存在 $r > 0, \theta_r(\cdot)$ 在 u 处取最小值". 对相反方向的结论, $v(\cdot)$ 的凸性是本质性的. 例如, 考虑函数 $v(u) := -|u|, u \in \mathbb{R}, u = 0$.

这一性质可推广到局部凸的拓扑向量空间的框架. 桶集 M 联系着相应的 Minkowski 度规函数 $p_M(\cdot)$, 见 (2.6).

命题 2.150 设 X 是局部凸的拓扑向量空间, 设 φ 是凸的, 值 $\mathrm{val}(\mathrm{P}_u)$ 有限. 则 (P_u) 是平静的当且仅当存在桶集 M, 它具有非空内部, 满足函数 $\theta_M(\cdot) := v(\cdot)+p_M(\cdot-u)$ 在点 u 处取最小值.

证明 设 M 是桶集, 满足相应的函数 $\theta_M(\cdot)$ 在 u 处取最小值. 有 $\theta_M(\cdot)$ 是凸的, 且

$$0 \leqslant \theta'_M(u,\cdot) = v'(u,\cdot)+p_M(\cdot),$$

因此对任何 $d\in M$, 有 $u'(u,d)\geqslant -1$. 由引理 2.8, 因为 M 具有非空内部且 $p_M(0)=0$, 有 $0\in \mathrm{int}(M)$. 这得到 $v'(u,\cdot)$ 在 0 的一邻域内是下方有界的, 由命题 2.126(i), $\partial v(u)$ 非空.

相反地, 设 $\partial v(u)$ 包含 u^*. 由于 u^* 是连续线性形式, 邻域基本系统具有形式 (2.10), 存在具有非空内部的桶集 M 满足 $|\langle u^*,d\rangle|\leqslant 1, \forall d\in M$, 因此 $|\langle u^*,d\rangle|\leqslant p_M(d)$. 置 $\theta_M(\cdot):=v(\cdot)+p_M(\cdot-u)$. 则对 $\forall u'\in U$,

$$\theta_M(u')\geqslant v(u)+\langle u^*,u'-u\rangle+p_M(u'-u)\geqslant v(u)=\theta_M(u),$$

证得结论. □

到目前为止用到的正则性条件不能推出最优值函数 $v(\cdot)$ 的连续性. 若假设 $v(\cdot)$ 在某一点 $\bar{u}$ 处是连续的, 则对相应的一对对偶问题有更强的结果.

定理 2.151 (对偶性定理) 设函数 $\varphi(x,u)$ 是正常的凸函数, 最优值函数 $v(u)=\mathrm{val}(\mathrm{P}_u)$ 是有限的, 在 $\bar{u}\in U$ 处连续. 则 $(\mathrm{P}_{\bar{u}})$ 与 $(\mathrm{D}_{\bar{u}})$ 间不存在对偶间隙, 集合 $\mathcal{S}(\mathrm{D}_{\bar{u}})$ 非空且等于 $\partial v(\bar{u})$, 且 $[\mathcal{S}(\mathrm{D}_{\bar{u}})]^\infty=\{0\}$. 再设 U 是赋予强拓扑的 Banach 空间. 则 $\mathcal{S}(\mathrm{D}_{\bar{u}})$ 是 U^* 的非空的凸的有界的弱 * 紧致的子集. 进一步, 最优值函数 $v(u)$ 在 $\bar{u}$ 处是 Hadamard 方向可微的, 且

$$v'(\bar{u},d)=\sup_{u^*\in\mathcal{S}(\mathrm{D}_{\bar{u}})}\langle u^*,d\rangle. \tag{2.278}$$

证明 由于 $v(u)$ 是凸的, 在 $\bar{u}$ 处连续, 由命题 2.126(iv) 得, $\partial v(\bar{u})$ 是非空的. 结合定理 2.142(i), 可推出第一与第二结论. 进一步有 $T_{\mathrm{dom}\ v}(\bar{u})=U$, 由命题 2.126(iii), $\partial v(\bar{u})$ 的回收锥是 $\{0\}$.

若还有 U 是 Banach 空间, 则由命题 2.126(v), $\partial v(\bar{u})$ 是 U^* 中凸的有界的弱 * 紧致子集合, $v(u)$ 在 $\bar{u}$ 处是 Hadamard 方向可微的, 其方向导数由集 $\partial v(\bar{u})$ 的支撑函数来计算. 这就证得结论. □

条件 $v(u)$ 在 $\bar{u}$ 处是连续的, 可视为约束规范. 它可以写为不同的等价形式. 例如, 若 $v(\cdot)$ 是凸的, $\mathrm{val}(\mathrm{P}_{\bar{u}})$ 是有限的, 由命题 2.108, 它等价于条件 $\mathrm{val}(\mathrm{P}_u)$ 在 $\bar{u}$ 的

邻域上是上方有界的. 进一步, 若空间 U 是有限维的, 则 $v(u)$ 在 $\bar{u}$ 处是连续的当且仅当 $\bar{u} \in \text{int}(\text{dom } v)$, 即对 $\bar{u}$ 的一邻域中的所有的 u, (P_u) 的可行集非空. 在无穷维的情形, 情况就不好处理. 下述结果是命题 2.111 的推广.

命题 2.152 设 X 与 U 是 Banach 空间. 设函数 $\varphi(x,u)$ 是正常的凸的下半连续的, 且 $v(\bar{u})$ 是有限的. 则 $v(u)$ 在 $\bar{u}$ 处是连续的当且仅当对 $\bar{u}$ 的一个邻域中的所有 u, 有 $v(u) < +\infty$, 即 $\bar{u} \in \text{int}(\text{dom } v)$.

证明 上述条件的必要性是显然的. 所以只需要证明充分性. 为此, 由命题 2.107, 只要证 $v(u)$ 在 $\bar{u}$ 的邻域上是上方有界的. 考虑如下定义的多值函数 $\mathcal{M}: X \times \mathbb{R} \to 2^U$,

$$\mathcal{M}(x,c) := \{u \in U : \varphi(x,u) \leqslant c\}.$$

因为 $\mathcal{M}$ 的图与 φ 的上图相同, φ 是凸的且闭的, 得 $\mathcal{M}$ 是凸的且闭的. 因为对 $\bar{u}$ 的邻域中的每一 u, 均有 $v(u) < +\infty$, 得到 $\bar{u} \in \text{int}(\text{range } \mathcal{M})$. 由于 $v(\bar{u})$ 是有限的, 存在一点 $(\bar{x},\bar{c})$ 满足 $\bar{u} \in \mathcal{M}(\bar{x},\bar{c})$. 由广义开映射定理2.70 得, 存在 $r > 0, \bar{u} \in \text{int}[\mathcal{M}(B_{X\times\mathbb{R}}((\bar{x},\bar{c}),r))]$①. 这推出, 对 $\bar{u}$ 的邻域中的所有点 u, 存在 x 满足 $\|x-\bar{x}\| < r$ 且 $\varphi(x,u) < \bar{c}+r$. 得到 $v(u)$ 在 $\bar{u}$ 的邻域上是上方有界的. 证毕. □

由上述命题易得下述结果.

命题 2.153 设 X 与 U 是 Banach 空间. 设函数 $\varphi(x,u)$ 是正常的凸的下半连续的, 且 $v(\bar{u})$ 有限, $\bar{u} \in \text{ri}(\text{dom } v)$. 则 $(\text{P}_{\bar{u}})$ 是平静的.

证明 不失一般性. 设 $\bar{u} = 0$. 令 $L := \text{cl}(\text{Sp}(\text{dom } v))$ 为由集 dom v 生成的线性闭空间. 因为 $0 \in \text{ri}(\text{dom } v)$, 将命题 2.152 限制到那一空间得最优值函数在 $0 \in L$ 处是连续的, 因而是次可微的, 即存在 $u^* \in L^*$ 满足对所有的 $u \in L$, $v(u) \geqslant v(0)+\langle u^*,u\rangle$. 由命题 2.11, 可将 u^* 延拓到 U^*. 因为, 若 $u \notin L$ 有 $v(u) = +\infty$, 这一延拓给出 v 在 0 处的一次梯度. □

例 2.154 令 $x = (x_1,x_2) \in \mathbb{R}^2$, $u \in \mathbb{R}$,

$$\varphi(x,u) := \begin{cases} x_1, & \text{若 } -x_1 + e^{x_2} + u \leqslant 0, \\ +\infty, & \text{否则}. \end{cases}$$

对任何 $\bar{u} \in \mathbb{R}$, 定理 2.151 中的假设是成立的, $v(u) = u, \forall u \in \mathbb{R}$. 还有

$$\varphi^*(0,u^*) = v^*(u^*) = \begin{cases} 0, & \text{若 } u^* = 1, \\ +\infty, & \text{否则}. \end{cases}$$

因此, (P_u) 与 (D_u) 间的对偶间隙是零, $\mathcal{S}(\text{D}_u) = \{1\}$. 有趣的是, 此例中, 原始问题 (P_u) 没有最优解, 但公式 (2.278) 成立.

① 原著中为 $\mathcal{M}(B_{X\times\mathbb{R}}(\bar{x},\bar{c}))$.

命题 2.155 设

(i) φ 是正常的下半连续的凸的函数.

(ii) 对给定 u, $\text{val}(\text{P}_u)$ 是有限的.

(iii) 对偶问题的最优解集$\mathcal{S}(\text{D}_u)$ 是非空的, 其回收锥是 $\{0\}$.

(iv) 集合 $\text{dom}\, v$ 具有非空的内部.

则 $u \in \text{int}(\text{dom}\, v)$; 若还有 X 和 U 是 Banach 空间, 则 $v(\cdot)$ 在 u 处是连续的, $\text{val}(\text{P}_u) = \text{val}(\text{D}_u)$, $\mathcal{S}(\text{D}_u)$ 是有界的.

证明 因为 $\text{val}(\text{D}_u) = v^{**}(u)$ 且 $\mathcal{S}(\text{D}_u)$ 非空, 有 $v^{**}(u)$ 是有限的. 由命题 2.141 得 $\mathcal{S}(\text{D}_u) = \partial v^{**}(u)$, 因而 $\partial v^{**}(u)$ 非空且其回收锥是 $\{0\}$. 由命题 2.126(iii), 这表明 $N_{\text{dom}\, v^{**}}(u) = \{0\}$. 由于 $\text{dom}\, v^{**}$ 与 $\text{dom}\, v$ 的拓扑包相同, 则由 $\text{dom}\, v$ 生成的线性空间在 U 中是稠密的. 因为设 $\text{dom}\, v$ 具有非空的相对内部, 得 $\text{dom}\, v$ 的内部是非空的. 结果由第一分离定理2.13, u 可从 $\text{dom}\, v$ 分离开来当且仅当 $u \notin \text{int}(\text{dom}\, v)$. 显然, 若 u 可以从 $\text{dom}\, v$ 分离, 则 $N_{\text{dom}\, v}(u)$, 从而 $N_{\text{dom}\, v^{**}}(u)$ 不是 $\{0\}$. 因此得到 $u \in \text{int}(\text{dom}\, v)$.

若还有 X 与 U 是 Banach 空间, 由命题 2.152, 由 $u \in \text{int}(\text{dom}\, v)$ 得 $v(\cdot)$ 在 u 处是连续的, 因此, 由定理 2.151 有 $\text{val}(\text{P}_u) = \text{val}(\text{D}_u)$, 且集合 $\mathcal{S}(\text{D}_u)$ 是有界的. □

若 U 是有限维的, U 中的每一凸集具有非空的相对内部. 此种情形, 上述命题中, $\text{dom}\, v$ 具有非空相对内部这一假设是多余的.

注意到, 对偶问题(D) 当然等价于

$$\min_{u^* \in U^*} \varphi^*(0, u^*). \tag{2.279}$$

这一问题可被嵌入到参数类优化问题

$$(\text{D}^{x^*}) \qquad \min_{u^* \in U^*} \varphi^*(x^*, u^*). \tag{2.280}$$

这一问题的对偶被称为 (共轭) 双对偶 (bidual) 问题. 若表示为极小化问题, 双对偶问题为

$$(\text{P}^{x^*}) \qquad \min_{x \in X} \{\varphi^{**}(x, 0) - \langle x^*, x\rangle\}. \tag{2.281}$$

对 $x^* = 0$, 问题 (D^0) 与对偶问题(2.279) 相同. 由 Fenchel-Moreau-Rockafellar 定理 2.113, 若 φ 是凸的闭的, 则 $\varphi^{**} = \varphi$, 因此非扰动的 (即 $x^* = 0$) 双对偶问题与原始问题 (P) 是相同的. 这种方式, (P) 与 (D) 间存在完整的对称关系. 另一方面, 给出最优值函数连续性, 因而次可微性的充分条件的命题 2.152 是基于 Banach 空间结构的. 因此, 对非自反的 Banach 空间, 为得到原始解的存在性是不可以将这一命题应用于对偶问题的.

考虑最优值函数

$$w(x^*) := \inf_{u^* \in U^*} \varphi^*(x^*, u^*), \tag{2.282}$$

即 (D^{x^*}) 的最优值. 设 φ 是凸的正常的下半连续的, 因此, $\varphi^{**} = \varphi$. 由命题 2.141 得 $\mathcal{S}(\mathrm{P}) = \partial w^{**}(0)$, 因此, (P) 有最优解当且仅当函数 $w^{**}(x^*)$ 在 $x^* = 0$ 处是次可微的. 进一步, 若问题 (P) 与 (D) 间没有对偶间隙, 则 $\partial w^{**}(0) = \partial w(0)$, 因此, 此种情况 (P) 具有最优解当且仅当最优值函数 $w(x^*)$ 在 $x^* = 0$ 处是次可微的. 在 X 与 U 是自反 Banach 空间的情形, 若 val(D) 是有限的, $w(x^*) < +\infty$ 对所有的接近 $0 \in X^*$ 的 x^* 均成立, 则 $w(\cdot)$ 在 $0 \in X^*$ 处是连续的 (见命题 2.152), 因而 (P) 具有非空有界的最优解集.

这一节引入了消除问题 (P_u) 与 (D_u) 的对偶间隙的三个较强的基本条件, 即用关于相应的最优值函数 $v(\cdot)$ 在 u 处的下半连续性、次可微性与连续性的条件.

2.5.2 Lagrange 对偶性

与 2.5.1 节中的参数或扰动形式的对偶相比较, 还有大家更为熟悉的 (min-max) 极小–极大或 Lagrange 对偶. 设 $K_X \subset X, K_Y \subset Y$ 是任意非空集合. 与函数 $\mathbb{L} : K_X \times K_Y \to \overline{\mathbb{R}}$ 相联系的原始与对偶问题分别定义为

$$\min_{x \in K_X} \sup_{y \in K_Y} \mathbb{L}(x, y), \tag{P^L}$$

$$\max_{y \in K_Y} \inf_{x \in K_X} \mathbb{L}(x, y). \tag{D^L}$$

称 $\mathbb{L}$ 是与上述问题相联系的极小–极大对偶性 Lagrange 函数(min-max duality Lagrangian). 差值 $\mathrm{val}(\mathrm{P}^L) - \mathrm{val}(\mathrm{D}^L)$(当它有意义时, 即 $\mathrm{val}(\mathrm{P}^L)$ 与 $\mathrm{val}(\mathrm{D}^L)$ 不取同样的无穷大值) 称为上述一对对偶问题的对偶间隙. 称 $(\bar{x}, \bar{y}) \in K_X \times K_Y$ 是函数 $\mathbb{L}(x, y)$ 的鞍点(saddle point), 若 $\mathbb{L}(\bar{x}, \bar{y}) \in \mathbb{R}$ 且

$$\mathbb{L}(\bar{x}, y) \leqslant \mathbb{L}(\bar{x}, \bar{y}) \leqslant \mathbb{L}(x, \bar{y}), \quad \forall (x, y) \in K_X \times K_Y. \tag{2.283}$$

命题 2.156　(i) *总成立着 $\mathrm{val}(\mathrm{D}^L) \leqslant \mathrm{val}(\mathrm{P}^L)$, 从而对偶间隙 $\mathrm{val}(\mathrm{P}^L) - \mathrm{val}(\mathrm{D}^L)$ 当它有意义时是非负值的.*

(ii) *函数 $\mathbb{L}(x, y)$ 有鞍点的充分必要条件是问题 (D^L) 与 (P^L) 具有相同的最优值, 且每一问题具有非空的最优解集. 此种情形, 鞍点集由 $\mathcal{S}(\mathrm{P}^L) \times \mathcal{S}(\mathrm{D}^L)$ 给出.*

证明　设 $(\hat{x}, \hat{y}) \in K_X \times K_Y$. 则

$$\inf_{x \in K_X} \mathbb{L}(x, \hat{y}) \leqslant \mathbb{L}(\hat{x}, \hat{y}) \leqslant \sup_{y \in K_Y} \mathbb{L}(\hat{x}, y),$$

因而有

$$\sup_{y\in K_Y}\inf_{x\in K_X}\mathbb{L}(x,y)\leqslant\inf_{x\in K_X}\sup_{y\in K_Y}\mathbb{L}(x,y).$$

于是得到 $\mathrm{val}(\mathrm{D}^L)\leqslant\mathrm{val}(\mathrm{P}^L)$, 对偶间隙是非负的.

现在设存在鞍点 $(\bar{x},\bar{y})$. 即有

$$\sup_{y\in K_Y}\mathbb{L}(\bar{x},y)\leqslant\mathbb{L}(\bar{x},\bar{y})\leqslant\inf_{x\in K_X}\mathbb{L}(x,\bar{y}).$$

上述不等式实际上成等式, 因为上确界与下确界分别在 $\bar{y}$ 与 $\bar{x}$ 处取得, 因而

$$\mathrm{val}(\mathrm{P}^L)\leqslant\sup_{y\in K_Y}\mathbb{L}(\bar{x},y)=\mathbb{L}(\bar{x},\bar{y})=\inf_{x\in K_X}\mathbb{L}(x,\bar{y})\leqslant\mathrm{val}(\mathrm{D}^L).$$

而 $\mathrm{val}(\mathrm{D}^L)\leqslant\mathrm{val}(P^L)$. 因而有 $\mathrm{val}(D^L)=\mathbb{L}(\bar{x},\bar{y})=\mathrm{val}(\mathrm{P}^L)$. 尤其由 $\mathrm{val}(\mathrm{P}^L)=\mathbb{L}(\bar{x},\bar{y})=\sup\limits_{y\in K_Y}\mathbb{L}(\bar{x},y)$, 得到 $\bar{x}\in\mathcal{S}(\mathrm{P}^L)$, 同样地, $\bar{y}\in\mathcal{S}(\mathrm{D}^L)$.

我们验证, 若原始最优值与对偶最优值相等, $\bar{x}\in\mathcal{S}(\mathrm{P}^L)$, $\bar{y}\in\mathcal{S}(\mathrm{D}^L)$, 则 $(\bar{x},\ \bar{y})$ 是 $\mathbb{L}$ 的鞍点. 这表明这一条件是充分的且鞍点集即 $\mathcal{S}(\mathrm{P}^L)\times\mathcal{S}(\mathrm{D}^L)$. 因为

$$\mathrm{val}(\mathrm{D}_L)=\inf_{x\in K_X}\mathbb{L}(x,\bar{y})\leqslant\mathbb{L}(\bar{x},\bar{y})\leqslant\sup_{y\in K_Y}\mathbb{L}(\bar{x},y)=\mathrm{val}(\mathrm{P}_L),$$

且原始与对偶最优值相等, 有

$$\mathbb{L}(\bar{x},\bar{y})=\inf_{x\in K_X}\mathbb{L}(x,\bar{y})\leqslant\mathbb{L}(x,\bar{y}),\quad\forall x\in K_X,$$

另一鞍点不等式可类似地证明. □

以下设 K_X 与 K_Y 分别是 Banach 空间 X 与 Y 的凸子集, 函数 $\mathbb{L}(x,y)$ 是方向可微的. 称 $(\bar{x},\bar{y})\in K_X\times K_Y$ 是 $\mathbb{L}(x,y)$ 关于 (K_X,K_Y) 的驻点 (critical point), 若 $\mathbb{L}(\bar{x},\bar{y})$ 是有限的, 且对 $\mathbb{L}_{\bar{y}}(\cdot):=\mathbb{L}(\cdot,\bar{y})$ 与 $\mathbb{L}_{\bar{x}}(\cdot):=\mathbb{L}(\bar{x},\cdot)$,

$$\mathbb{L}'_{\bar{y}}(\bar{x},h)\geqslant 0\ \text{且}\ \mathbb{L}'_{\bar{x}}(\bar{y},d)\leqslant 0,\quad\forall(h,d)\in\mathcal{R}_{K_X}(\bar{x})\times\mathcal{R}_{K_Y}(\bar{y}).\tag{2.284}$$

称 $\mathbb{L}(x,y)$ 是凸–凹的 (convex-concave), 若对任意的 $(\hat{x},\hat{y})\in K_X\times K_Y$, 函数 $\mathbb{L}(\cdot,\hat{y})$ 与 $\mathbb{L}(\hat{x},\cdot)$ 分别是凸的和凹的.

命题 2.157 若 K_X 与 K_Y 是凸集合, 则 $\mathbb{L}$ 的鞍点集被包含在 $\mathbb{L}$ 的驻点集中. 若 $\mathbb{L}$ 是凸–凹的, 则这两个集合相同.

证明 设 $(\bar{x},\bar{y})$ 是 $\mathbb{L}$ 的鞍点, $(h,d)\in\mathcal{R}_{K_X}(\bar{x})\times\mathcal{R}_{K_Y}(\bar{y})$. 因为 K_X 与 K_Y 是凸集, 对 $t>0$ 充分小有 $\bar{x}+th\in K_X$, $\bar{y}+td\in K_Y$. 因此

$$\frac{\mathbb{L}(\bar{x},\bar{y}+td)-\mathbb{L}(\bar{x},\bar{y})}{t}\leqslant 0\leqslant\frac{\mathbb{L}(\bar{x}+th,\bar{y})-\mathbb{L}(\bar{x},\bar{y})}{t}.$$

令 $t \downarrow 0$, 得到驻点的刻画 (2.284).

若 $(\bar{x},\bar{y})$ 是 $\mathbb{L}$ 的驻点, 且 $\mathbb{L}$ 是凸–凹的, 用凸函数(凹函数) 大于 (小于) 方向导数 的性质, 对所有的 $(x,y) \in K_X \times K_Y$, 有

$$\mathbb{L}(\bar{x},y) \leqslant \mathbb{L}(\bar{x},\bar{y}) + \mathbb{L}'_{\bar{x}}(\bar{y},y-\bar{y}) \leqslant \mathbb{L}(\bar{x},\bar{y}) \leqslant \mathbb{L}(\bar{x},\bar{y}) + \mathbb{L}'_{\bar{y}}(\bar{x},x-\bar{x}) \leqslant \mathbb{L}(x,\bar{y}),$$

证得 $(\bar{x},\bar{y})$ 是 $\mathbb{L}$ 的鞍点. □

为看 min-max 对偶性与共轭对偶性间的关系, 考虑下述函数, 它可视为与 (P_u) 及其 (共轭) 对偶 (D_u) 相联系的对偶性 (duality)Lagrange 函数

$$\mathcal{L}(x,u^*,u) := \langle u^*,u\rangle - \varphi_u^*(x,u^*), \tag{2.285}$$

其中, 对给定的 x, φ_u^* 是关于 u 的偏共轭:

$$\varphi_u^*(x,u^*) = \sup_{u'\in U}\{\langle u^*,u'\rangle - \varphi(x,u')\}.$$

则得到

$$\inf_{x\in X}\mathcal{L}(x,u^*,u) = \langle u^*,u\rangle - \sup_{(x,u')\in X\times U}\{\langle u^*,u'\rangle - \varphi(x,u')\},$$

从而

$$\inf_{x\in X}\mathcal{L}(x,u^*,u) = \langle u^*,u\rangle - \varphi^*(0,u^*).$$

因此, 对偶问题(D_u) 等价于

$$\max_{u^*\in U^*}\left\{\inf_{x\in X}\ \mathcal{L}(x,u^*,u)\right\}. \tag{2.286}$$

另一方面, 得到

$$\sup_{u^*\in U^*}\mathcal{L}(x,u^*,u) = \sup_{u^*\in U^*}\left\{\langle u^*,u\rangle - \varphi_u^*(x,u^*)\right\} = \varphi_u^{**}(x,u). \tag{2.287}$$

因此, 由 $\varphi_u^{**} \leqslant \varphi$ 得

$$\mathrm{val}(\mathrm{D}_u) = \sup_{u^*\in U^*}\inf_{x\in X}\mathcal{L}(x,u^*,u) \leqslant \inf_{x\in X}\sup_{u^*\in U^*}\ \mathcal{L}(x,u^*,u) = \inf_{x\in X}\varphi_u^{**}(x,u) \leqslant \mathrm{val}(\mathrm{P}_u).$$

进一步, 若对 $\forall x \in X$, 函数 $\varphi(x,\cdot)$ 是凸的且闭的, 则由 Fenchel-Moreau-Rockafellar 定理得

$$\sup_{u^*\in U^*}\ \mathcal{L}(x,u^*,u) = \varphi(x,u), \tag{2.288}$$

则原始问题 (P_u) 可表示为下述形式

$$\min_{x\in X}\left\{\sup_{u^*\in U^*}\mathcal{L}(x,u^*,u)\right\}. \tag{2.289}$$

若 $\varphi(x,\cdot)$ 是正常的凸的闭的函数, 则 (P_u) 与 (D_u) 可以通过交换应用到对偶性 Lagrange 函数 $\mathcal{L}(x,u^*,u)$ 取极大 (max) 与取极小 (min) 算子得到. 即此种情形下共轭对偶性与极小-极大对偶性是重合的, 且若 $\bar{x}$, $\bar{u}^*$ 分别是原始问题与对偶问题的最优解, 这两个问题的最优值相等, 则 $(\bar{x},\ \bar{u}^*)$ 是 $\mathcal{L}(\cdot,\cdot,u)$ 在 $X\times U^*$ 的鞍点, 即

$$\sup_{u^*\in U^*}\mathcal{L}(\bar{x},u^*,u)=\mathcal{L}(\bar{x},\bar{u}^*,u)=\inf_{x\in X}\mathcal{L}(x,\bar{u}^*,u). \tag{2.290}$$

注意到, 后面在 (2.296) 中定义的 "标准 Lagrange 函数", 不要同极小-极大对偶性 Lagrange 函数或对偶 Lagrange 函数相混淆. 除非额外说明, Lagrange 函数指的是标准的 Lagrange 函数.

2.5.3 对偶理论的例子与应用

设 X, X^*, 与 Y, Y^* 是成对的局部凸的拓扑向量空间. 下述参数 (共轭) 对偶性的例子在后续的讨论中尤其重要. 考虑优化问题

$$(\mathrm{P})\qquad \min_{x\in X}\{f(x)+F(G(x))\}, \tag{2.291}$$

其中 $f:X\to\overline{\mathbb{R}}, F:Y\to\overline{\mathbb{R}}$ 是正常函数, $G:X\to Y$. 上述问题 (P) 的可行域为

$$\Phi:=\{x\in\mathrm{dom}\ f:G(x)\in\mathrm{dom}\ F\}. \tag{2.292}$$

注意到, 若 $F(\cdot):=I_K(\cdot)$ 是非空集 $K\subset Y$ 的指示函数, 则问题 (P) 具有下述形式

$$(\mathrm{P})\qquad \min_{x\in X}f(x)\quad \text{s.t.}\quad G(x)\in K. \tag{2.293}$$

后面将讨论这一特殊情况.

将 (P) 嵌入到下述问题族中

$$(\mathrm{P}_y)\qquad \min_{x\in X}\{f(x)+F(G(x)+y)\}, \tag{2.294}$$

其中 $y\in Y$ 视为参数向量. 显然, $y=0$ 时, 相应的问题 (P_0) 即与 (P) 重合. 问题 (P_y) 是下述函数关于 x 的极小化问题

$$\varphi(x,y)=f(x)+F(G(x)+y). \tag{2.295}$$

记 $v(y)$ 为相应的最优值函数, 即 $v(y)=\mathrm{val}(\mathrm{P}_y)$, 或等价地, $v(y)=\inf_{x\in X}\varphi(x,y)$.

注意到, 函数 $\varphi(x,u)$ 的定义域是非空的, 事实上, 因为 f 与 F 是正常的, 则存在 $x\in\mathrm{dom}(f)$ 及 $y\in\mathrm{dom}(F)$. 则

$$\varphi(x,y-G(x))=f(x)+F(y)<+\infty,$$

因而 $(x, y - G(x)) \in \text{dom}(\varphi)$. 进一步, 对 $\forall (x,y) \in X \times Y$, $\varphi(x,y) \geqslant -\infty$, 因而 φ 是正常的. 若 f 与 F 是下半连续的且 G 是连续的, 则函数 φ 是下半连续的. 尤其, 若 $F(\cdot) := I_K(\cdot)$ 是指示函数, 则它是下半连续的当且仅当集合 K 是闭的.

令

$$L(x, y^*) := f(x) + \langle y^*, G(x)\rangle \tag{2.296}$$

为问题 (P) 的标准 Lagrange 函数(standard Lagrangian)(后面就称为 Lagrange 函数). 有

$$\begin{aligned}\varphi^*(x^*, y^*) &= \sup_{x\in X,\ y\in Y}\{\langle x^*, x\rangle + \langle y^*, y\rangle - f(x) - F(G(x)+y)\}\\ &= \sup_{x\in X}\{\langle x^*, x\rangle - f(x) - \langle y^*, G(x)\rangle\\ &\quad + \sup_{y\in Y}[\langle y^*, G(x)+y\rangle - F(G(x)+y)]\}.\end{aligned}$$

在上述等式中的右端的第二个上确界中作变量替换 $G(x) + y \to y$, 得到

$$\varphi^*(x^*, y^*) = \sup_{x\in X}\{\langle x^*, x\rangle - L(x, y^*)\} + F^*(y^*).$$

因此, (共轭) 对偶问题(D_y) 可以写为如下形式 (对共轭对偶的一般性的定义, 见 (2.266))

$$(\text{D}_y) \qquad \max_{y^*\in Y^*}\{\langle y^*, y\rangle + \inf_{x\in X} L(x, y^*) - F^*(y^*)\}. \tag{2.297}$$

尤其, 对 $y = 0$, (P) 的对偶是

$$(\text{D}) \qquad \max_{y^*\in Y^*}\{\inf_{x\in X} L(x, y^*) - F^*(y^*)\}. \tag{2.298}$$

因 $\text{val(P)} \geqslant \text{val(D)}$, 若对 $x_0 \in X$, $\bar{y}^* \in Y^*$, 原始与对偶目标函数值相等, 即

$$f(x_0) + F(G(x_0)) = \inf_x L(x, \bar{y}^*) - F^*(\bar{y}^*), \tag{2.299}$$

则 $\text{val(P)} = \text{val(D)}$, 若其公共值是有限的, 则 $x_0 \in X$ 与 $\bar{y}^* \in Y^*$ 分别是 (P) 与 (D) 的最优解. 条件 (2.299) 可以写为下述等价形式

$$(L(x_0, \bar{y}^*) - \inf_x L(x, \bar{y}^*)) + (F(G(x_0)) + F^*(\bar{y}^*) - \langle \bar{y}^*, G(x_0)\rangle) = 0. \tag{2.300}$$

显然, (2.300) 的左端的第一项非负, 由 Young-Fenchel 不等式(2.213) 得第二项也是非负的. 进一步, 等式

$$F(G(x_0)) + F^*(\bar{y}^*) - \langle \bar{y}^*, G(x_0)\rangle = 0$$

成立当且仅当 $\bar{y}^* \in \partial F(G(x_0))$. 所以, 条件 (2.299) 等价于

$$x_0 \in \underset{x \in X}{\operatorname{argmin}}\, L(x, \bar{y}^*) \text{ 且 } \bar{y}^* \in \partial F(G(x_0)). \tag{2.301}$$

结合定理 2.142, 得到下述结果.

定理 2.158 若 $\operatorname{val}(P) = \operatorname{val}(D)$, $x_0 \in X$, $\bar{y}^* \in Y^*$ 分别是 (P) 与 (D) 的最优解, 则条件 (2.301) 成立. 相反地, 若条件 (2.301) 对点 x_0 与 $\bar{y}^*$ 成立, 则 x_0 是 (P) 的最优解, $\bar{y}^*$ 是 (D) 的最优解, (P) 与 (D) 之间没有对偶间隙.

注 2.159 设 f 是方向可微的, G 是 Gâteaux 可微的. 令 $(x_0, \bar{y}^*)$ 满足 (2.301). 则 $L(\cdot, \bar{y}^*)$ 是方向可微的, (2.301) 的第一条件表明, $L(\cdot, \bar{y}^*)$ 在 x_0 的方向导数均是非负的, 即

$$f'(x_0, h) + \langle \bar{y}^*, DG(x_0)h \rangle \geqslant 0, \quad h \in X.$$

若还有 f 是 Gâteaux 可微的, 则由于 $\bar{y}^* \in \partial F(G(x_0))$, 得到

$$0 \in Df(x_0) + [DG(x_0)]^* \partial F(G(x_0)). \tag{2.302}$$

类似地, 若 f 是凸的, 则 $0 \in \partial f(x_0) + [DG(x_0)]^* \partial F(G(x_0))$.

假设 $F(\cdot)$ 是正常的凸的下半连续函数, 可以从 2.5.2 节引入的 Lagrange 对偶的角度导出对偶问题(D_y). 对任意 $x \in X$, 函数 $\varphi(x, \cdot) = f(x) + F(G(x) + \cdot)$ 是正常的凸的下半连续的. 由 2.5.2 节中的讨论, 通过定义对偶性 Lagrange 函数

$$\mathcal{L}(x, y^*, y) = \langle y^*, y \rangle + L(x, y^*) - F^*(y^*), \tag{2.303}$$

极小-极大对偶性等价于共轭性对偶.

如果最优值函数 $v(y)$ 在 $y = 0$ 处是次可微的, 则问题 (P) 是平静的. 由命题 2.147 得, 若 (P) 是平静的, 则 $\operatorname{val}(P) = \operatorname{val}(D)$ 且对偶问题的最优解集非空.

命题 2.160 设最优值函数 $v(\cdot)$ 是凸的, $\operatorname{val}(P)$ 是有限的, 或者空间 Y 是有限维的, 或者函数 $F(\cdot)$ 在点 $\bar{y} \in Y$ 的邻域上是上方有界的. 则由 (2.291) 给出的问题 (P) 是平静的当且仅当下述条件成立

$$\liminf_{t \downarrow 0} \frac{v(td) - v(0)}{t} > -\infty, \quad \forall d \in Y. \tag{2.304}$$

证明 若空间 Y 是有限维的, 则由命题 2.134 得, 条件 (2.304) 是 $v(y)$ 在 $y = 0$ 处的次可微性的充分必要条件. 每一情况, 条件 (2.304) 是 $v(y)$ 在 $y = 0$ 处的次可微性的必要条件. 现在设 $F(\cdot)$ 在一点 $\bar{y} \in Y$ 的邻域上是上方有界的, 令 $\bar{x} \in \operatorname{dom} f$. 考虑点 $y_0 := \bar{y} - G(\bar{x})$. 对 $\forall y$,

$$v(y) \leqslant f(\bar{x}) + F(G(\bar{x}) + y).$$

因为对 y_0 的邻域上的所有的 y, $F(G(\bar{x})+y)$ 是上方有界的, $v(\cdot)$ 在 y_0 的这一邻域上是上方有界的. 由命题 2.134 得 $v(y)$ 在 $y=0$ 处是次可微的当且仅当条件 (2.304) 成立. □

定义 2.161　称由 (2.291) 给出的问题 (P) 是凸的, 若函数 $F(\cdot)$ 是下半连续的, 函数 $f(x)$ 与 $\psi(x,y):=F(G(x)+y)$ 是凸的.

若问题 (P) 是凸的, 则 $\psi(x,y)=f(x)+F(G(x)+y)$ 是凸的, 因而最优值函数 $v(y)$ 也是凸的. 显然, 由 $F(G(x)+y)$ 的凸性可推出函数 $F(\cdot)$ 与 $F(G(\cdot))$ 的凸性. 相反的结论一般来说是不真的, 即 $F(y)$ 与 $F(G(x))$ 的凸性不一定推出函数 $F(G(x)+y)$ 的凸性. 例如, 取 $F(y):=|y|$ 及 $G(x)=x^2$, $x,y\in\mathbb{R}$.

命题 2.162　设函数 $F(\cdot)$ 是凸的. 则函数 $\psi(x,y):=F(G(x)+y)$ 是凸函数的充要条件是映射 $\mathcal{G}(x,c):=(G(x),c):X\times\mathbb{R}\to Y\times\mathbb{R}$ 关于集合 $(-\text{epi }F)$ 是凸的.

证明　由定义, 映射 $\mathcal{G}$ 关于集合 $(-\text{epi }F)$ 是凸的当且仅当多值函数 $\mathcal{M}(x,c):=(G(x),c)-\text{epi }F$ 是凸的. 有

$$\begin{aligned}\text{gph }\mathcal{M}&=\{(x,c_1,y,c_2):(G(x),c_1)-(y,c_2)\in\text{epi }F\}\\&=\{(x,c_1,y,c_2):F(G(x)-y)\leqslant c_1-c_2\}.\end{aligned}$$

显然, 函数 $\psi(x,y)$ 是凸的当且仅当函数 $\phi(x,y):=F(G(x)-y)$ 是凸的. 显然

$$\text{epi }\phi=\{(x,y,c):F(G(x)-y)\leqslant c\}.$$

因此, $\mathcal{M}$ 的图是凸的当且仅当 ϕ 的上图是凸的. 从而多值函数 $\mathcal{M}$ 是凸的当且仅当函数 $\phi(x,y)$ 是凸的. 这就完成了证明. □

现在考虑情况 $F(\cdot):=I_K(\cdot)$, 其中 K 是 Y 的非空闭凸子集. 此种情况下 (P) 可以写为 (2.293) 的形式. 因为 $I_K^*(y^*)=\sigma(y^*,K)$, 则相应的对偶问题变为

$$\text{(D)}\qquad \max_{y^*\in Y^*}\{\inf_{x\in X}L(x,y^*)-\sigma(y^*,K)\}. \tag{2.305}$$

定义 2.163　称具有形式 (2.293) 的问题 (P) 是凸的, 若函数 $f(x)$ 是凸的, 集合 K 是闭的凸的, 映射 $G(x)$ 相应于集合 $C:=-K$ 是凸的.

由命题 2.162 得, 上述定义等价于复合函数 (2.291) 由 2.161 定义的凸性. 即 $G(x)$ 关于 (凸) 集 $-K$ 是凸的当且仅当函数 $\psi(x,y):=I_K(G(x)+y)$ 是凸的. 这是因为函数 $I_K(G(x)-y)$ 的上图即多值函数 $\mathcal{M}(x):=G(x)-K$ 的图与 $\mathbb{R}_+$ 的积. 由 (P) 的凸性可推出 $\varphi(x,y)=f(x)+I_K(G(x)+y)$ 的凸性, 从而得到最优值函数 $v(y):=\inf\limits_{x\in X}\varphi(x,y)$ 的凸性.

注意到, $I_K(\cdot)$ 在某一点的邻域上是上方有界的充要条件是 K 具有非空内部. 因此, 由命 2.160 可得下述结果.

命题 2.164 *考虑具有形式 (2.293) 的问题 (P). 设问题 (P) 是凸的, val(P) 是有限的, 或者 Y 是有限维的, 或者 K 具有非空内部. 则 (P) 是平静的充要条件是条件 (2.304) 成立.*

注意到, 由于 $\partial I_K(y_0) = N_K(y_0)$, 最优条件 (2.301) 可表示为

$$x_0 \in \operatorname*{argmin}_{x \in X} L(x, \bar{y}^*) \text{ 且 } \bar{y}^* \in N_K(G(x_0)). \tag{2.306}$$

若集合 K 是闭凸锥 (convex cone), 则

$$\sup_{y^* \in K^-} L(x, y^*) = \begin{cases} f(x), & \text{若 } G(x) \in K, \\ +\infty, & \text{若 } G(x) \notin K. \end{cases}$$

因此, 此种情况, 原始问题具有下述形式

$$\text{(P)} \qquad \min_{x \in X} \sup_{y^* \in K^-} L(x, y^*). \tag{2.307}$$

此种情况, 若 $y^* \in K^-$, $\sigma(y^*, K) = 0$, 否则 $\sigma(y^*, K) = +\infty$, 因此对偶问题具有下述形式

$$\text{(D)} \qquad \max_{y^* \in K^-} \inf_{x \in X} L(x, y^*). \tag{2.308}$$

因此, 对于锥约束, 原始与对偶问题可以通过将 “max” 与 “min” 运算应用于 Lagrange 函数 $L(x, y^*)$, 限定 y^* 属于 K^-, 交换它们的顺序得到, 总成立着 (见 (2.268)) $\operatorname{val}(P) \geqslant \operatorname{val}(D)$. 因此, 若 $\operatorname{val}(P) = -\infty$, 即原始问题 (P) 是下方无界的, 则 $\operatorname{val}(D) = -\infty$.

注意到, 由 (2.306) 的第二条件得 $G(x_0) \in K$(因为否则法锥$N_K(G(x_0))$ 是空集), 即 x_0 是 (P) 的可行点, 且 $\bar{y}^* \in (K^\infty)^-$. 进一步, 若 (P) 是凸的, 因此 (见 2.3.5 节) G 关于锥 $(-K^\infty)$ 是凸的, 得 $\langle \bar{y}^*, G(\cdot) \rangle$ 是凸函数, 因此 $L(\cdot, \bar{y}^*)$ 是凸的. 若还有 $L(\cdot, \bar{y}^*)$ 是 (Gâteaux) 可微的, 则 x_0 是 $L(\cdot, \bar{y}^*)$ 的极小点当且仅当 $D_x L(x_0, \bar{y}^*) = 0$. 结果, 若 (P) 是凸的, f 与 G 是可微的, 则条件 (2.306) 可以写为下述的等价形式

$$D_x L(x_0, \bar{y}^*) = 0 \text{ 且 } \bar{y}^* \in N_K(G(x_0)). \tag{2.309}$$

再注意到, 若 K 是凸锥, 则条件 $\bar{y}^* \in N_K(G(x_0))$ 等价于

$$G(x_0) \in K, \ \bar{y}^* \in K^- \text{ 且 } \langle \bar{y}^*, G(x_0) \rangle = 0. \tag{2.310}$$

现在考虑命题 2.152 的正则性条件, 这要求在 0 的邻域上所有的 y 有 $v(y) < +\infty$, 或等价地 $0 \in \operatorname{int}(\operatorname{dom} v)$, 有 $v(y) < +\infty$ 当且仅当 (P_y) 具有可行点 x 满足 $f(x) < +\infty$, 即存在 $x \in \operatorname{dom} f$ 满足 $G(x) + y \in K$, 即

$$\operatorname{dom} v = K - G(\operatorname{dom} f). \tag{2.311}$$

所以, 在此种情形, 正则性条件 $0 \in \text{int}(\text{dom}\, v)$ 可以表示为 (与 2.3.4 节中的约束规范 (2.202) 相比较)

$$0 \in \text{int}\{G(\text{dom}\, f) - K\} \tag{2.312}$$

(注意, 类似的结果可用于 (2.291) 形式的问题 (P), 其中 K 被替换为集合 $\text{dom}\, F$). 尤其, 条件 (2.312) 可推出可行集 Φ 是非空的, 从而 $\varphi(x,y)$ 是正常的. 若还有 $f(x)$ 是下半连续的且 $G(x)$ 是连续的, 则 $\varphi(x,y)$ 是下半连续的. 设 X 与 Y 是 Banach 空间, 问题 (P) 的最优值 $\text{val(P)} = v(0)$ 是有限的. 由命题 2.152 得 $v(\cdot)$ 在 Y 的零点的邻域内是连续的. 因此, 由命题 2.148 与对偶性定理 2.151, 得到下述结果.

定理 2.165　*设 X 与 Y 是 Banach 空间, 问题 (P) 具有形式 (2.293). 设 $f(x)$ 是下半连续的, $G(x)$ 是连续的, 问题 (P) 是凸的且满足正则性条件 (2.312). 则问题 (P) 与 (D) 间没有对偶间隙, 即 $\text{val(P)} = \text{val(D)}$. 进一步, 若 (P) 的最优值是有限的, 则对偶问题(D) 的最优解集是 Y^* 的非空的凸的有界的弱 * 紧致子集.*

下述结果是将命题 2.155 用于此种情况的特殊情形.

命题 2.166　*设*

(i) 具有形式 (2.293) 的问题 (P) 是凸的, $f(x)$ 是下半连续的, $G(x)$ 是连续的.

(ii) val(P) 是有限的.

(iii) 对偶问题的最优解集 $\mathcal{S}(\text{D})$ 是非空的, 其回收锥是 $\{0\}$.

(iv) 集合 $G(\text{dom}\, f) - K$ 具有非空的相对内部.

*则正则性条件 (2.312) 成立, 若 X 与 Y 是 Banach 空间, 则 $\text{val(P)} = \text{val(D)}$ 且 $\mathcal{S}(\text{D})$ 是有界的弱 * 紧致的.*

注意到, 若空间 Y 是有限维的或 K 具有非空的内部, 则上述条件 (iv) 是成立的. 下述例子表明, 上述命题中这一条件是本质性的.

例 2.167　令对所有的 $x \in X$, $f(x) = 0$, $K = \{0\}$, $G : X \to Y$ 是连续的线性算子满足 $G(X) \neq Y$ 且 $G(X)$ 在 Y 中是稠密的(即 $\text{cl}[G(X)] = Y$). 例如, 可取 $X = C[0,1]$, $Y = L_2[0,1]$, 定义 G 为 $C[0,1]$ 到 $L_2[0,1]$ 的自然嵌入, 即 $G(x)$ 是 $L_2[0,1]$ 中对应于连续函数 $x(\cdot) \in C[0,1]$ 的等价类. 显然, 对应最优化问题 (P) 具有有限的最优值 $\text{val(P)} = 0$. 也有

$$v(y) = \begin{cases} 0, & \text{若 } -y \in G(X), \\ +\infty, & \text{否则}. \end{cases}$$

因此, 函数 $v(y)$ 在 $y = 0$ 处不是连续的, 且正则性条件 (2.312) 不成立. 另一方面

$$v^*(y^*) = \begin{cases} 0, & \text{若 } y^* = 0, \\ +\infty, & \text{若 } y^* \neq 0, \end{cases}$$

且 $\partial v(0) = \{0\}$, 因此 $\text{val(P)} = \text{val(D)}$, (D) 的最优解集是非空有界的.

设 $f(x)$ 是下半连续的且 $G(x)$ 是连续的, 问题 (P) 是凸的, 因此, 集合 $Q :=$ dom f 是凸的闭的. 由命题 2.104, 若还有映射 $G(x)$ 是连续可微的, 则条件 (2.312) 等价于在每一可行点 $x \in \Phi$ 的 Robinson 约束规范

$$0 \in \text{int}\{G(x) + DG(x)(Q - x) - K\}. \tag{2.313}$$

(广义)Slater 条件对问题 (P) 成立, 若存在 $\bar{x} \in \text{dom}\, f$ 满足 $G(\bar{x}) \in \text{int}\, K$(见 2.3.4 节). 显然 Slater 条件推出正则条件 (2.312). 若 K 是闭凸集, 具有非空内部, 相反的结论也成立 (见命题 2.106).

2.5.4 应用于次微分理论

这一节讨论对偶理论到次微分理论的应用, 其目的在于研究复合函数的次微分与分量的微分性质之间的关系. 令 $f : X \to \overline{\mathbb{R}}$, $g : Y \to \overline{\mathbb{R}}$ 是正常的凸函数, $A : X \to Y$ 是连续的线性映射. 考虑复合函数

$$F(x) := f(x) + g(Ax),$$

其定义域为

$$\text{dom}\, F = \{x \in \text{dom}\, f : Ax \in \text{dom}\, g\}.$$

考虑极小化 F 的下述问题

$$\text{(P)} \qquad \min_{x \in X}\{f(x) + g(Ax)\}. \tag{2.314}$$

此问题可视为问题 (2.291) 的特殊情形. 这一问题的 Lagrange 函数是 $L(x, y^*) = f(x) + \langle A^*y^*, x\rangle$, 因而

$$\inf_{x \in X} L(x, y^*) = -\sup_{x \in X}\{-f(x) + \langle -A^*y^*, x\rangle\} = -f^*(-A^*y^*).$$

因此, 相联系的对偶问题(见 (2.298)) 可表示为

$$\text{(D)} \qquad \max_{y^* \in Y^*}\{-f^*(-A^*y^*) - g^*(y^*)\}. \tag{2.315}$$

注意到这两个对偶问题的对称性. 也注意到与问题 (P) 相联系的相应的最优值函数的定义域 dom v 由下述给出

$$\text{dom}\, v = \text{dom}\, g - A(\text{dom}\, f), \tag{2.316}$$

因而正则性条件 $0 \in \text{int}(\text{dom}\, v)$ 可以表示为

$$0 \in \text{int}\{A(\text{dom}\, f) - \text{dom}\, g\} \tag{2.317}$$

(见 (2.311), (2.312) 及所述).

定理 2.168 (Moreau-Rockafellar)　设 X 与 Y 是 Banach 空间, $f : X \to \overline{\mathbb{R}}$, $g : Y \to \overline{\mathbb{R}}$ 是正常的 l.s.c. 凸函数, $A : X \to Y$ 是连续的线性映射, $F(x) := f(x)+g(Ax)$. 设正则条件 (2.317) 成立. 则对任何 $x_0 \in \mathrm{dom}\ F$, 有

$$\partial F(x_0) = \partial f(x_0) + A^*[\partial g(Ax_0)]. \tag{2.318}$$

证明　设 $x^* \in \partial f(x_0)$, $y^* \in \partial g(Ax_0)$. 则对 $\forall h \in X$, $f(x_0+h) \geqslant f(x_0)+\langle x^*, h\rangle$, $g(A(x_0+h)) \geqslant g(Ax_0) + \langle y^*, Ah\rangle$, 将这两个不等式相加得到

$$F(x_0+h) \geqslant F(x_0) + \langle x^* + A^*y^*, h\rangle,$$

因而得到 $x^* + A^*y^* \in \partial F(x_0)$. 这证得 (2.318) 的右端包含在 $\partial F(x_0)$.

相反地, 令 $q \in \partial F(x_0)$. 这意味着 $F(x) - \langle q, x\rangle$ 在 x_0 处取得最小, 即 x_0 是下述问题的最优解

$$\min_{x\in X}\{F(x) - \langle q, x\rangle\}.$$

置 $\tilde{f}(x) = F(x) - \langle q, x\rangle$. 由于当 f 变到 $\tilde{f}$ 时, 正则性条件 (2.317) 是不变的, 则存在对偶问题的解 y^*, y^* 满足对应着 (2.301) 的最优性条件, 即 $y^* \in \partial g(Ax_0)$ 且 $0 \in \partial f(x_0) - q + A^*y^*$. 或等价地, 存在 $x^* \in \partial f(x_0)$ 及 $y^* \in \partial g(Ax_0)$, $q = x^* + A^*y^*$. 这证得结论. □

注 2.169　尤其, 对 $X = Y$, 单位映射 A, 我们得到, 若 $f, g : X \to \overline{\mathbb{R}}$ 是正常的下半连续的凸函数且正则条件

$$0 \in \mathrm{int}\{\mathrm{dom}\ f - \mathrm{dom}\ g\} \tag{2.319}$$

成立, 则对任何 $x_0 \in (\mathrm{dom}\ f) \cap (\mathrm{dom}\ g)$, 得到

$$\partial(f+g)(x_0) = \partial f(x_0) + \partial g(x_0). \tag{2.320}$$

注意到, 要求 f 与 g 是下半连续的, 这是保证 (2.320) 成立的本质性条件. 例如, 取线性非连续函数 $f : X \to \overline{\mathbb{R}}$ 与 $g := -f$, 则对任意 $x \in X$, 有

$$\partial f(x) = \partial g(x) = \varnothing, \quad 而\ f(\cdot) + g(\cdot) = 0, \quad \partial(f+g)(x) = \{0\}.$$

注 2.170　正则性条件 (2.317) 成立的一个有用的充分条件是存在 $x_0 \in \mathrm{dom}(f)$, 满足 g 在 Ax_0 处是连续的.

由 (2.315) 给出的对偶问题(D) 可嵌入到下述参数族中

$$(\mathrm{D}^{x^*}) \qquad \max_{y^*\in Y^*}\{-f^*(x^* - A^*y^*) - g^*(y^*)\}. \tag{2.321}$$

设 f 与 g 是正常的 l.s.c. 的凸函数, 有 $f^{**}=f$, $g^{**}=g$. 则这一问题的对偶, 即 (2.314) 所述的问题 (P) 的双对偶是

$$(\mathrm{P}^{x^*}) \qquad \min_{x\in X}\{f(x)-\langle x^*,x\rangle+g(Ax)\}, \tag{2.322}$$

它的最优值为 $-(f+g\circ A)^*(x^*)$. 对 $x^*=0$, 其双对偶问题与原始问题 (P) 是重合的, 即对偶扰动归功于目标函数 f 的线性扰动 (这是 (2.280) 之后所讨论情况的一个特例). 因为 (原始) 正则条件 (2.317) 在此 (对偶) 扰动下具有不变性, 由定理 2.168 得, 若条件 (2.317) 成立, 则对所有的 $x^*\in X^*$,

$$(f+g\circ A)^*(x^*)=\inf_{y^*\in Y^*}\{f^*(x^*-Ay^*)+g^*(y^*)\}, \tag{2.323}$$

上述等式的右端的下确界在 Y^* 的非空弱 * 紧致子集上是可达到的 (上面等式的左端和右端分别等于问题 (2.322) 的最优值的负值与问题 (2.321) 的最优值的负值).

一个有意思的特殊情况是 $X=Y$, 且 A 是单位映射, 它与应用到函数 ψ, $\phi: X\to\overline{\mathbb{R}}$ 的所谓下卷积算子(infimal convolution operator) 相联系：

$$\psi\diamond\phi(u):=\inf_{x\in X}\{\psi(u-x)+\phi(x)\}. \tag{2.324}$$

容易验证, $\diamond$ 是变换的 (commutative) 结合的 (associative) 算子, 即 $\psi\diamond\phi=\phi\diamond\psi$ 与 $(\psi_1\diamond\psi_2)\diamond\psi_3=\psi_1\diamond(\psi_2\diamond\psi_3)$. 因为 $\psi\diamond\phi$ 是函数 $\varphi(x,u):=\psi(u-x)+\phi(x)$ 关于 x 的极小化问题的最优值函数, 我们可用对偶理论.

定理 2.171 设 X 是 Banach 空间, f, $g: X\to\overline{\mathbb{R}}$ 是正常的 l.s.c. 凸函数, 设正则条件 (2.319) 成立. 则

$$(f+g)^*=f^*\diamond g^*, \tag{2.325}$$

如果 $\bar{x}^*$ 满足 $(f+g)^*(\bar{x}^*)\in\mathbb{R}$, 则满足下式的 x^* 的集合

$$(f^*\diamond g^*)(\bar{x}^*)=f^*(\bar{x}^*-x^*)+g^*(x^*)$$

是非空的且弱 * 紧致的.

注 2.172 令 X 与 Y 是 Banach 空间. 由命题 2.153 知, 比 (2.317) 弱的依然能保证 $\mathrm{val(P)}=\mathrm{val(D)}$ 且 $\mathcal{S}(D)$ 非空的条件是

$$0\in\mathrm{ri}\{A(\mathrm{dom}\,f)-\mathrm{dom}\,g\}. \tag{2.326}$$

类似地, 较 (2.319) 弱能保证 (2.325) 成立的条件是

$$0\in\mathrm{ri}\{\mathrm{dom}\,f-\mathrm{dom}\,g\}. \tag{2.327}$$

后面需要线性映射 $A \in \mathcal{L}(X,Y)$ 与其伴随 (adjoint mapping) 映射 $A^*: Y^* \to X^* \in \mathcal{L}(Y^*, X^*)$ 的下述性质.

命题 2.173　令 X 与 Y 是 Banach 空间, $A: X \to Y$ 是连续的线性映射, $A^*: Y^* \to X^*$ 是其伴随映射. 则

(i) $(\ker A)^\perp = \mathrm{cl}^*(\mathrm{range}(A^*))$.

(ii) 若 range (A) 是闭的, 则 $(\ker A)^\perp = \mathrm{range}(A^*)$, 且存在 $c > 0$ 满足对所有 $x^* \in \mathrm{range}(A^*)$, 存在 $y^* \in Y^*$, 有 $\|y^*\| \leqslant c\|x^*\|$ 与 $x^* = A^*y^*$.

(iii) 若还有, $\mathrm{range}(A) = Y$, 即 A 是映上的, 则 A^* 是一对一的, 且存在 $c > 0$, 满足对所有的 $y^* \in Y^*$, 有 $\|y^*\| \leqslant c\|A^*y^*\|$.

(iv) $(\ker A^*)^\perp = \mathrm{cl}(\mathrm{range}(A))$.

证明　(i) 令 $y^* \in Y^*$ 且 $x^* = A^*y^*$. 则对任何 $x \in \ker A$, 有 $\langle x^*, x\rangle = \langle y^*, Ax\rangle = 0$, 因而有 $x^* \in (\ker A)^\perp$. 由于 $(\ker A)^\perp$ 是弱 * 闭集合, 于是得到 $\mathrm{cl}^*(\mathrm{range}(A^*)) \subset (\ker A)^\perp$.

现在证相反的包含关系. 设 x^* 属于 $(\ker A)^\perp$, 但不属于 $\mathrm{cl}^*(\mathrm{range}(A^*))$. 则由第二分离定理 2.14, 存在 $x \in X$ 满足

$$\langle x^*, x\rangle < \inf\{\langle \hat{x}^*, x\rangle : \hat{x}^* \in \mathrm{range}(A^*)\}.$$

因为 $\mathrm{range}(A^*)$ 是向量空间, 对任何 $\hat{x} \in \mathrm{range}(A^*)$, $\langle \hat{x}^*, x\rangle = 0$, 因此对任何 $y^* \in Y^*$, 有 $\langle y^*, Ax\rangle = \langle A^*y^*, x\rangle = 0$. 结果 $x \in \ker A$ 且 $\langle x^*, x\rangle < 0$, 这与 x^* 的假设是矛盾的. 证得 (i).

(ii) 令 $c \in (\ker A)^\perp$, 考虑最优化问题

$$\min_x \langle c, x\rangle \qquad \text{s.t.} \qquad Ax + y = 0,$$

其中 $y \in Y$ 是参数. 相联系的最优值函数 $v(y)$ 满足 $v(0) = 0$ 且定义域是 $\mathrm{range}(A)$, 设为是闭的. 由命题 2.153 得非扰动问题 $(y = 0)$ 是平静的, 即对偶问题至少有一最优解. 由于对偶问题是在 $c + A^*\lambda = 0, \lambda \in Y^*$ 的约束下, 求零目标函数的极大值, 这可推出 $c \in \mathrm{range}(A^*)$. 于是得到 $(\ker A)^\perp \subset \mathrm{range}(A^*)$. 结合 (i) 得到 $(\ker A)^\perp = \mathrm{range}(A^*)$.

置 $Y_1 := \mathrm{range(A)}$. 由开映射定理, 存在 $\varepsilon > 0$ 满足 $\varepsilon B_{Y_1} \subset A(B_X)$. 所以

$$\|A^*y^*\| = \sup_{x \in B_X} \langle y^*, Ax\rangle \geqslant \varepsilon \sup_{y \in B_{Y_1}} \langle y^*, y\rangle = \varepsilon \|y_1^*\|,$$

其中 y_1^* 是 y^* 到 Y_1 的限制. 由 Hahn-Banach 定理, 存在 $\tilde{y}^* \in Y^*$ 满足 $\|\tilde{y}^*\| = \|y_1^*\|$ 且 $\forall y_1 \in Y_1$ 有 $\langle \tilde{y}^*, y_1\rangle = \langle y^*, y_1\rangle$①, 得到 $x^* = A^*y^*$ 满足 $x^* = A^*\tilde{y}^*$, $\|\tilde{y}^*\| \leqslant \varepsilon^{-1}\|x^*\|$.

① 原书中这里的 $\tilde{y}^*$ 为 $\tilde{y}$.

(iii) 若 $\text{range}(A) = Y$, 由结论 (ii) 的上述证明, 存在 $\varepsilon > 0$, 对任意的 $y^* \in Y^*$, $\|A^*y^*\| \geqslant \varepsilon\|y_1^*\|$. 于是得到 (iii).

(iv) 设 $x_0 \in X$, 置 $y_0 := Ax_0$. 则对任意的 $y^* \in \ker A^*$, 有 $\langle y^*, y_0\rangle = \langle y^*, Ax_0\rangle = \langle A^*y^*, x_0\rangle = 0$, 因此有 $y_0 \in (\ker A^*)^\perp$. 因为 $(\ker A^*)^\perp$ 是闭集, 得到 $\text{cl}(\text{range}(A)) \subset (\ker A^*)^\perp$.

设 $y_1 \notin \text{cl}(\text{range}(A))$. 则由第二分离定理 2.14, 存在 $y^* \in Y^*$ 满足

$$\langle y^*, y_1\rangle < \inf\{\langle y^*, y\rangle : y \in \text{range}(A)\}.$$

由 $\text{range}(A)$ 是向量空间得, 对 $\forall y \in \text{range}(A)$ 有 $\langle y^*, y\rangle = 0$, 因此有 $A^*y^* = 0$. 由于上面的下确界是 0, $\langle y^*, y_1\rangle < 0$, 因而得 $y_1 \notin (\ker A^*)^\perp$. 这证得 (iv). □

2.5.5 紧致集上最大值函数的极小化

设 X 是 Banach 空间, Ω 是非空的紧致的度量空间, Q 是 X 的非空的凸的闭子集, $g : X \times \Omega \to \mathbb{R}$ 是连续函数. 考虑最优化问题

$$(\text{P}') \qquad \min_{x\in Q} \sup_{\omega\in\Omega} g(x, \omega). \tag{2.328}$$

取 $f(\cdot) := I_Q(\cdot)$, $G(x)(\cdot) := g(x, \cdot)$, $F(y) := \sup\limits_{\omega\in\Omega} y(\omega)$, 则问题 (P') 可视为由 (2.291) 定义的问题的特殊情况. 注意到, 函数 $F : C(\Omega) \to \mathbb{R}$ 是连续的凸的正齐次的, 且 $F(\cdot) = \sigma(\cdot, \partial F(0))$.

下述命题表明, 映射 $G(x)(\cdot) := g(x, \cdot)$ 把点 $x \in X$ 映到函数 $g(x, \cdot) \in C(\Omega)$, 且继承了函数 $g(x, \omega)$ 的许多性质.

命题 2.174 设函数 $g(x, \omega)$ 以 $X \times \Omega$ 的乘积拓扑关于联合变量 x 与 ω 连续. 则下述结论成立:

(i) 映射 $G : X \to C(\Omega)$ 是连续的.

(ii) 对每一 $\omega \in \Omega$, 函数 $g(\cdot, \omega)$ 是 (Gâteaux) 可微的且 $D_x g(x, \omega)$ 关于 x 与 ω 联合变量是连续的, 则 $G(x)$ 是连续可微的, 且

$$[DG(x)h](\cdot) = D_x g(x, \cdot)h. \tag{2.329}$$

(iii) 映射 $G(x)$ 关于非负值函数构成的锥 $C_+(\Omega)$ 是凸的当且仅当函数 $g(\cdot, \omega)$ 对所有的 $\omega \in \Omega$ 是凸函数.

证明 (i) 考虑点 $x_0 \in X_0$. 因为 $g(x, \omega)$ 是连续的, Ω 是紧致的, 对任何 $\varepsilon > 0$, 存在 x_0 的邻域 N_x 和有限个点 $\omega_i \in \Omega$, $i = 1, \cdots, k$, 及相应的邻域 N_i, 满足 $\Omega = \bigcup\limits_{i=1}^{k} N_i$ 及对所有的 $i = 1, \cdots, k$,

$$|g(x, \omega) - g(x_0, \omega_i)| \leqslant \varepsilon, \quad \forall (x, \omega) \in N_x \times N_i.$$

则对 $\forall x \in N_x$ 及 $\forall \omega \in \Omega$,

$$|g(x,\omega) - g(x_0,\omega)| \leqslant 2\varepsilon, \quad \forall \omega \in \Omega,$$

这表明 $G(\cdot)$ 在点 x_0 处是连续的.

(ii) 考虑 $x,\ h \in X$ 及 (固定)$\omega \in \Omega$. 由中值定理得

$$g(x+h,\omega) - g(x,\omega) = \int_0^1 D_x g(x+th,\omega)h\mathrm{d}t,$$

因此得

$$|g(x+h,\omega) - g(x,\omega) - D_x g(x,\omega)h| \leqslant \|h\| \int_0^1 \|D_x g(x+th,\omega) - D_x g(x,\omega)\| \,\mathrm{d}t.$$

由 $D_x g(x,\omega)$ 的连续性及 Ω 的紧致性得

$$\sup_{t\in[0,1]} \|D_x g(x+th,\omega) - D_x g(x,\omega)\|$$

当 $h \to 0$ 时关于 $\omega \in \Omega$ 一致收敛到 0. 于是得 $G(\cdot)$ 在 x 处是可微的且公式 (2.329) 成立. $G(x)$ 的连续可微性由 $D_x g(x,\omega)$ 与 Ω 的紧致性得到.

结论 (iii) 由定义立即得到. □

$C(\Omega)$ 的对偶空间由 Ω 的有限符号的 Borel 测度构成. 对测度 $\mu \in C(\Omega)^*$, 考虑积分函数 $\gamma(x) := \int_\Omega g(x,\omega)\ \mathrm{d}\mu(\omega)$.

命题 2.175 设函数 $g(x,\omega)$ 关于联合变量 x 与 ω 是连续的, 对所有的 $\omega \in \Omega$, $g(\cdot,\omega)$ 是凸函数. 令 $\mu \in C(\Omega)^*$ 是非负测度, $\gamma(x) := \int_\Omega g(x,\omega)\mathrm{d}\mu(\omega)$ 是相应的积分函数. 则

(i) $\gamma(x)$ 是凸的连续的, 且对 $x,\ h \in X$,

$$\gamma'(x,h) = \int_\Omega g'_\omega(x,h)\mathrm{d}\mu(\omega), \tag{2.330}$$

其中 $g'_\omega(x,h)$ 记 $g(\cdot,\omega)$ 在 x 处沿方向 h 的方向导数.

(ii) 若进一步, $g(\cdot,\omega)$ 在 x 处对 μ 几乎处处的 ω 是 Gâteaux 可微的, 则 $\gamma(\cdot)$ 在 x 处是 Hadamard 可微的且

$$D\gamma(x)h = \int_\Omega D_x g(x,\omega)h\mathrm{d}\mu(\omega). \tag{2.331}$$

证明 (i) 若测度 μ 是离散的, 即 $\mu = \sum_{i=1}^m a_i\delta(\omega_i)$, 则 $\gamma(x) = \sum_{i=1}^m a_i g(x,\omega_i)$, 因为 μ 是非负的, 因此 $a_i \geqslant 0$, $i = i,\cdots,m$, $g(\cdot,\omega_i)$ 是凸函数, 得到 $\gamma(x)$ 是凸的. 由于凸函数序列的逐点极限是凸的, 对一般测度 μ, 由极限结论, 得 $\gamma(x)$ 的凸性.

考虑点 $x \in X$. 由于 Ω 是紧致的, $g(x,\omega)$ 是连续的, 得到存在常数 c 及 x 的邻域 N_x 满足对所有 $x \in N_x$ 与 $\omega \in \Omega$ 有 $g(x,\omega) \leqslant c$. 得到 $\gamma(\cdot)$ 在 N_x 上是上方有界的, 界值是 $c\mu(\Omega)$, 因此它在 x 处是连续的 (见命题 2.107).

考虑收敛到零的单调递减正数序列 $\{t_n\}$ 及相联系的函数序列

$$\psi_n(\omega) := \frac{g(x+t_nh,\omega)-g(x,\omega)}{t_n}.$$

由 $g(\cdot,\omega)$ 的凸性, 对任意固定 $\omega \in \Omega$, 序列 $\{\psi_n(\omega)\}$ 是单调递减的, 当 $n \to \infty$ 时收敛到 $g'_\omega(x,h)$. 由单调收敛定理 (monotone convergence theorem) 得

$$\lim_{n\to\infty} \frac{\gamma(x+t_nh)-\gamma(x)}{t_n} = \lim_{n\to\infty} \int_\Omega \psi_n(\omega)\mathrm{d}\mu(\omega) = \int_\Omega g'_\omega(x,h)\mathrm{d}\mu(\omega),$$

即公式 (2.330) 成立. 注意到, 由于 $\gamma(\cdot)$ 是连续的, 因而在 x 的邻域内是 Lipschitz 连续, 则方向导数$\gamma'(x,\cdot)$ 是有限值的且连续的.

(ii) 若 $g(\cdot,\omega)$ 在 x 处是 Gâteaux 可微的, 有 $g'_\omega(x,h) = D_xg(x,\omega)h$. 则由 (2.330) 得, 若这一等式对 μ 几乎处处的 ω 是成立的, 则

$$\gamma'(x,h) = \int_\Omega D_xg(x,\omega)h\mathrm{d}\mu(\omega).$$

得到 $\gamma'(x,\cdot)$ 是线性的, 因为 $\gamma(\cdot)$ 是连续的, 则可得 $\gamma(\cdot)$ 的 Hadamard 可微性. □

在例子 2.122 中已证明 $\partial F(0) = \mathcal{P}_\Omega$, 因此 $F^*(\cdot) = I_{\mathcal{P}_\Omega}(\cdot)$, 其中

$$\mathcal{P}_\Omega := \{\mu \in C(\Omega)^* : \mu(\Omega) = 1, \quad \mu \succeq 0\}$$

记 Ω 上的概率测度的集合. 结果得到问题 (P′)(见 (2.298)) 的对偶

$$(\mathrm{D}') \qquad \max_{\mu\in\mathcal{P}_\Omega} \inf_{x\in Q} \int_\Omega g(x,\omega)\mathrm{d}\mu(\omega). \tag{2.332}$$

命题 2.176 *设 X 是 Banach 空间, $Q \subset X$ 是非空凸闭集, Ω 是紧致的度量空间, $g: X\times\Omega \to \mathbb{R}$ 是连续函数, 考虑由 (2.328) 与 (2.332) 给出的最优化问题 (P′) 与 (D′). 设对任何 $\omega \in \Omega$, 函数 $g(\cdot,\omega)$ 是凸的. 则*

(i) 问题 (P′) 与 (D′) 具有相同的最优值.

*(ii) 若 (P′) 与 (D′) 的公共的最优值是有限的, 则问题 (D′) 具有非空的凸的弱 * 紧致的最优解集.*

(iii) $x_0 \in X$ 与 $\bar\mu \in \mathcal{P}_\Omega$ 分别是 (P′) 与 (D′) 的最优解当且仅当

$$x_0 \in \operatorname*{argmin}_{x\in Q} \int_\Omega g(x,\omega)\mathrm{d}\bar\mu(\omega) \text{ 且 } \operatorname{supp}(\bar\mu) \subset \operatorname*{argmax}_{\omega\in\Omega} g(x_0,\omega). \tag{2.333}$$

(iv) 若还有 $g(\cdot,\omega)$ 在 x_0 处对 $\bar{\mu}$ 几乎处处的 ω 是 Gâteaux 可微的，则 $x_0 \in X$, $\bar{\mu} \in \mathcal{P}_\Omega$ 分别是 (P′) 与 (D′) 的最优解的充分必要条件是

$$-\int_\Omega D_x g(x_0,\omega)\mathrm{d}\bar{\mu}(\omega) \in N_Q(x_0) \text{ 且 } \operatorname{supp}(\bar{\mu}) \subset \operatorname*{argmax}_{\omega\in\Omega} g(x_0,\omega). \tag{2.334}$$

证明 考虑与问题 (P′) 相联系的最优值函数

$$v(y) := \inf_{x\in Q}\max_{\omega\in\Omega}\{g(x,\omega)+y(\omega)\}, \quad y \in C(\Omega),$$

得到 $\operatorname{dom} v = C(\Omega)$, 且对任何 $y \in C(\Omega)$ 有 $v(y) < +\infty$. 由命题 2.152 与定理 2.151 得 $\operatorname{val}(\mathrm{P}') = \operatorname{val}(\mathrm{D}')$, 且 (D′) 的最优解集是非空的凸的弱 * 紧致的. 由最大值函数次微分的公式 (2.236), 条件 (2.333) 是相应的条件 (2.301) 的特殊情况. 最后，若 $g(\cdot,\omega)$ 在 x_0 处对 $\bar{\mu}$ 几乎处处的 ω 是 Gâteaux 可微的，则相应的积分函数在 x_0 处是 (Hadamard) 可微的且公式 (2.331) 成立. 结果，此种情形，条件 (2.333) 与 (2.334) 是等价的. □

在无限维的空间 X 的情形，可证明，由 (2.332) 给出的对偶问题(D′) 有有限支撑的最优解测度. 更精确地，有下述结果.

命题 2.177 *设命题 2.176 的条件成立，$X = \mathbb{R}^n$, 设原始问题 (P′) 具有最优解是 x_0. 则相应的对偶问题 (D′) 有最优解 $\bar{\mu}$, 它的支撑 $\operatorname{supp}(\bar{\mu})$ 至多有 $n+1$ 点.*

证明 由命题 2.176 得 $\mathcal{S}(\mathrm{D}')$ 是非空的凸的弱 * 紧致的. 由 Krein-Milman 定理2.19, 它至少有一极点. 现在证 $\mathcal{S}(\mathrm{D}')$ 的任何极点有至多 $n+1$ 个点的支撑. 事实上，设 μ 是 $\mathcal{S}(\mathrm{D}')$ 的一个极点，设 μ 的支撑多于 $n+1$ 点. 则可将 $\operatorname{supp}(\mu)$ 划分成 $n+2$ 个非空的互不相交的 Borel 集 $\Omega_i, i=1,\cdots,n+2$, 满足 $\mu(\Omega_i)>0, i=1,\cdots,n+2$. 记 μ_i 是 μ 到 Ω_i 的限制，即 $\mu_i(A) := \mu(A\cap\Omega_i)$. 注意到 μ_i 是正的测度，且 $\mu = \sum_{i=1}^{n+2}\mu_i$. 函数 $\gamma_i(x) := \int_{\Omega_i} g(x,\omega)\mathrm{d}\mu(\omega)$ 是凸的连续的，且 $\sum_{i=1}^{n+2}\gamma_i(x) := \int_\Omega g(x,\omega)\mathrm{d}\mu(\omega)$. 由 (2.333) 得 x_0 是函数 $\sum_{i=1}^{n+2}\gamma_i(x)+I_Q(x)$ 的极小点，因此，由注 2.169, $0 \in \sum_{i=1}^{n+2}\partial\gamma_i(x_0)+\partial I_Q(x_0)$. 于是存在 $q_i \in \partial\gamma_i(x_0)$, $i=1,\cdots,n+2$ 满足 $-\sum_{i=1}^{n+2} q_i \in N_Q(x_0)$. 考虑含 $n+2$ 个未知数 $\alpha = (\alpha_1,\cdots,\alpha_{n+2})$ 的 $n+1$ 个方程的线性系统：

$$\sum_{i=1}^{n+2}\alpha_i q_i = 0, \quad \sum_{i=1}^{n+2}\alpha_i\mu(\Omega_i) = 0.$$

由于方程的个数小于未知数的个数，系统有非零解 $\bar{\alpha} \in \mathbb{R}^{n+2}$. 对充分小的 $\varepsilon > 0$, 所

有的数 $1\pm\varepsilon\bar{\alpha}_i,\ i=1,\cdots,n+2$ 是正的, 且

$$\mu=\frac{1}{2}\bigg(\sum_i(1-\varepsilon\bar{\alpha}_i)\mu_i+\sum_i(1+\varepsilon\bar{\alpha}_i)\mu_i\bigg).$$

这意味着 μ 是两个不同的概率测度的算术平均数, 每一测度满足对应 (2.333) 的最优性条件, 它们的支撑等于 μ 的支撑. 因此, μ 是 $\mathcal{S}(\mathrm{D}')$ 中两个不同元素的算术平均数, 这与 μ 的极点性质矛盾. □

考虑问题 (P') 的下述离散化

$$(\mathrm{P}'_m)\qquad \min_{x\in Q}\max_{i\in\{i,\cdots,m\}} g(x,\omega_i), \tag{2.335}$$

其中 $\omega_i,\cdots,\omega_m\in\Omega$ 是固定的点. 由有限维凸情形的命题 2.177 得, 总可以构造问题 (P') 的有限的离散化, 它具有相同的最优值与最优解. 更精确地, 下述结果成立.

命题 2.178 设 $X=\mathbb{R}^n$, $Q\subset X$ 是非空凸闭集, Ω 是紧致的度量空间, $g: X\times\Omega\to\mathbb{R}$ 是连续函数. 设对每一 $\omega\in\Omega$, 函数 $g(\cdot,\omega)$ 是凸的, 问题 (P') 有最优解. 则存在点 $\omega_1,\cdots,\omega_{n+1}\in\Omega$ 满足对应的离散化问题 (P'_{n+1}) 与问题 (P') 具有相同的最优值与相同的最优解集.

证明 设 x_0 是 (P') 的最优解. 由命题 2.177 得对偶问题(D') 具有最优解 $\bar{\mu}$, 它的支撑具有至多 $n+1$ 个点, 即存在点 $\omega_1,\cdots,\omega_{n+1}\in\Omega$ 满足存在 $\overline{\lambda}_i>0$, 使得 $\bar{\mu}=\sum_{i=1}^{n+1}\overline{\lambda}_i\delta(\omega_i)$. 考虑对应这些点的离散化问题 (P'_{n+1}). 问题 (P'_{n+1}) 的对偶问题可表示为

$$(\mathrm{D}'_{n+1})\qquad \max_{\lambda\in\Xi}\inf_{x\in Q}\bigg\{\sum_{i=1}^{n+1}\lambda_i g(x,\omega_i)\bigg\}, \tag{2.336}$$

其中

$$\Xi:=\bigg\{\lambda=(\lambda_1,\cdots,\lambda_{n+1}):\lambda\geqslant 0,\ i=1,\cdots,n+1,\ \sum_{i=1}^{n+1}\lambda_i=1\bigg\}.$$

显然, $\mathrm{val}(\mathrm{D}')\geqslant\mathrm{val}(\mathrm{D}'_{n+1})$. 有离散测度 $\bar{\mu}$ 满足最优性条件 (2.333). 因为这些最优性条件对相应的离散化问题也是收敛的, 所以, $x_0,\overline{\lambda}$ 分别是问题 (P'_{n+1}) 与 (D'_{n+1}) 的最优解. 于是得到 $\mathrm{val}(\mathrm{D}')=\mathrm{val}(\mathrm{D}'_{n+1})$. 由命题 2.176, $\mathrm{val}(\mathrm{D}')=\mathrm{val}(\mathrm{P}')$ 与 $\mathrm{val}(\mathrm{D}'_{n+1})=\mathrm{val}(\mathrm{P}'_{n+1})$. 最后, 若 $\hat{x}$ 是 (P'_{n+1}) 是最优解, 则与它对应着 (D'_{n+1}) 的最优解 $\hat{\lambda}$, 满足相应的最优性条件. 于是得到 $\hat{x}$ 与 $\hat{\mu}:=\sum_{i=1}^{n+1}\hat{\lambda}_i\delta(\omega_i)$ 分别是问题 (P') 与 (D') 的最优解, 完成命题的证明. □

由 2.5.3 节已经看到, 如果原始问题如下

(P)
$$\min_{x\in X} f(x)+F(G(x)),\tag{2.337}$$

其中 f 与 G 是连续函数, F 是正常的 l.s.c. 凸函数, 则相联系的对偶问题具有形式 (见 (2.298))

(D)
$$\max_{y^*\in Y^*}\inf_{x\in X} L(x,y^*)-F^*(y^*)^{①}\tag{2.338}$$

这等价于下述函数的极小化问题

$$\phi(y^*):=F^*(y^*)+\sup_{x\in X}\{-L(x,y^*)\}.\tag{2.339}$$

$L(\cdot,y^*)$ 在 X 上的 (可能空的) 闭子集的上下确界是可以取到的. 设 $\bar{y}^*$ 是对偶问题的最优解, 因而有 $0\in\partial\phi(\bar{y}^*)$, 且存在紧致集 $\Omega\subset X$ 满足对 $\bar{y}^*$ 的邻域中的 y^*, $L(\cdot,y^*)$ 在 Ω 的 (紧致的) 子集上的下确界是可取到的. 则 $y^*\mapsto\sup\limits_{x\in X}(-L(x,y^*))$ 是凸连续的, 因此在 $\bar{y}^*$ 处是次可微的. 由注 2.169 得

$$0\in\partial\phi(\bar{y}^*)=\partial\Big(\sup_{x\in X}(-L(x,y^*))\Big)+\partial F^*(\bar{y}^*\Big),$$

或等价地, 存在 $q\in\partial F^*(\bar{y}^*)$ 满足

$$0\in\partial\Big(\sup_{x\in X}(\langle q,x\rangle-L(x,y^*))\Big).$$

再由命题 2.176, 存在概率测度μ, 其支持在集 $\Omega^*:=\mathrm{argmin}\ L(\cdot,\bar{y}^*)$ 上, 满足

$$\int_{\Omega^*}(q-G(x))\mathrm{d}\mu(x)=0.$$

由这一关系推出 $\displaystyle\int_{\Omega^*}G(x)\mathrm{d}\mu(x)=q\in\partial F^*(\bar{y}^*)$, 因此

$$\bar{y}^*\in\partial F\left(\int_{\Omega^*}G(x)\mathrm{d}\mu(x)\right).\tag{2.340}$$

上述讨论是优化问题松弛 (relaxation of optimization problems) 理论的根源, 其中原始问题被下述松弛问题替换

(RP)
$$\min_{\mu\in\mathcal{P}_\Omega}\int_\Omega f(x)\mathrm{d}\mu(x)+F\left(\int_\Omega G(x)\mathrm{d}\mu(x)\right),\tag{2.341}$$

其中 Ω 是 X 的 “充分大” 的紧致子集. 由 2.5.3 节的结果, 这一凸问题的对偶是

$$\max_{y^*\in Y^*}\inf_{\mu\in\mathcal{P}_\Omega}\left(\int_\Omega L(x,y^*)\mathrm{d}\mu(x)-F^*(y^*)\right),\tag{2.342}$$

① 原著中为 min.

或等价地,

$$\max_{y^* \in Y^*} \inf_{x \in \Omega} (L(x, y^*) - F^*(y^*)). \tag{2.343}$$

注意到, 在我们的条件下, 上述问题的值等于 (D) 的值. 由上述讨论知道, 存在 $\mu \in \mathcal{P}_\Omega$, 它的支撑被包含在 $\operatorname*{argmin}_{x \in X} L(x, y^*)$, 且满足 (2.340). 因为这些关系可推出零对偶间隙且刻画 (RP) 与对偶 (见 (2.301)) 的解, 我们已经证得下述结论.

命题 2.179 设 (i) 对偶问题(D) 具有最优解 $\bar{y}^*$; (ii) 对充分接近 $\bar{y}^*$ 的 y^*, 集合 $\operatorname*{argmin}_{x \in X} L(x, y^*)$ 非空且被包含在紧致集 Ω 中. 则问题 (RP) 与 (P) 具有相同的最优值, 松弛问题(RP) 的最优解集非空, 且 $\mu \in \mathcal{S}(\mathrm{RP})$ 当且仅当 $\mathrm{supp}(\mu) \subset \operatorname*{argmin}_{x \in X} L(x, \bar{y}^*)$ 且 (2.340) 成立.

在上述命题的假设下, 若 (RP) 的解集包含支撑在单个点 x_0 上的测度, 则 $(x_0, \bar{y}^*)$ 满足联系着 (P) 与 (D) 的最优性系统, 因此 x_0 是 (P) 的最优解. 若 $\operatorname{argmin} L(\cdot, \bar{y}^*)$ 只含有一单个点, 这种情况当然就出现了. 我们得到由 (2.291) 与 (2.298) 给出的问题 (P) 与对偶问题(D) 的下述结果.

命题 2.180 设 $f : X \to \mathbb{R}$, $G : X \to Y$ 是连续函数, 设 F 是正常的 l.s.c. 凸函数, 并设 (i) 对偶问题(D) 有最优解 $\bar{y}^*$; (ii) $\operatorname*{argmin}_{x \in X} L(x, \bar{y}^*) = \{x_0\}$; (iii) 对充分接近于 $\bar{y}^*$ 的 y^*, 集合 $\operatorname*{argmin}_{x \in X} L(x, y^*)$ 是非空的, 且包含在紧致集 Ω 中. 则问题 (P) 与 (D) 具有相同的最优解, 且 x_0 是原始问题的最优解.

注意到, 在上述框架下, 若 $X = \mathbb{R}^n$, 则由命题 2.177 得, 对偶问题具有最优解测度, 它的支撑至多有 $n+1$ 个点.

作为研究极小–极大问题的结果, 我们得到著名的 Helly 定理(的变形).

命题 2.181 (Helly 定理) 设 A_i, $i \in I$ 是 (可能无限的)$\mathbb{R}^n$ 的一族闭凸子集. 设 A_i, $i \in I$ 没有公共的非零的回收方向, 且这类集合中的任何 $n+1$ 个集合之交非空. 则这类集合的交是非空的.

证明 由上述假设得, 这类集合的交 $\bigcap_{i \in I} A_i$ 是有界 (尽管可能是空) 的. 事实上, $\mathbb{R}^n$ 的凸子集是有界的充要条件是没有非零的回收方向. 所以, 若 $\bigcap_{i \in I} A_i$ 是无界的, 它含有一非零的回收方向. 结果, 所有的 A_i, $i \in I$ 会有公共的非零的回收方向. 所以, 存在 $r > 0$ 满足对所有的 $x \in \bigcap_{i \in I} A_i$ 有 $\|x\| \leqslant r$. 置 $\mathcal{Q} := \{x \in \mathbb{R}^n : \|x\| \leqslant r\}$.

下证集合族 A_i, $i \in I$ 的任何有限子类与集合 $\mathcal{Q}$ 具有非空的交. 事实上, 记 $d_i(x) := \inf_{y \in A_i} \|x - y\|$ 为到集合 A_i 的距离函数. 因为 A_i 是凸的, 这些函数是连续的凸的. 给定有限集 $J \subset I$, 考虑最优化问题:

$$(\mathrm{P}_J) \qquad \min_{x \in \mathcal{Q}} \{\max_{i \in J} d_i(x)\}. \tag{2.344}$$

上述问题是由 (2.328) 定义的问题 (P′) 的特殊情况, 其中集 Ω 由集 J 给出. 因为 $\mathcal{Q}$ 是紧致的, 距离函数是连续的, 因而函数 $\max\limits_{i\in J} d_i(x)$ 是连续的, 这一问题有 (可能非唯一的) 最优解, 记为 x_J. 由命题 2.177 得 (P_J) 的对偶具有最优解 λ, 至多有 $n+1$ 个非零的分量. 令 $J' := \{i \in J : \lambda_i \neq 0\}$. 对所有的 $i \in J'$ 有 $\mathrm{val}(\mathrm{P}_J) = d_i(x_J)$, x_J 满足问题 $(\mathrm{P}_{J'})$ 的最优性条件, $(\mathrm{P}_{J'})$ 的最优值等于 (P_J) 的最优值. 因为此问题是凸的, 所以下确界实际上在 x_J 处达到. 另一方面, 因为 $|J'| \leqslant n+1$, 由定理的假设有 $\bigcap\limits_{i\in J'} A_i$ 是非空的. 因为 $\bigcap\limits_{i\in J'} A_i \subset \bigcap\limits_{i\in I} A_i \subset \mathcal{Q}$, 这些问题的公共最优值是 0, 这意味着子类 $\{A_i,\ i \in J\}$ 具有非空的交.

我们证得集合类 $A_i' := \mathcal{Q} \cap A_i$ 满足任何有限子类均有非空的交, 因为这些集合是紧致集 $\mathcal{Q}$ 的子集, 由命题 2.4 得它们的交是非空的. □

在 $\mathbb{R}^n$ 的有限凸子集类的情形, 可以拿掉每一集合的闭性及交集的有界性这样的假设. 下述结果是经典的 Helly 定理.

命题 2.182 (Helly 定理)　设 $A_i,\ i \in I$ 是 $\mathbb{R}^n$ 的有限个 (不必是闭的) 凸子集的族. 设这一族集合中任何 $n+1$ 个集合的交是非空的. 则这一族集合的交是非空的.

证明　在集合族 $A_i,\ i \in I$ 中由 $n+1$ 个集合构成的子类中的交中取一点. 令 T 是所有这些点的集合. 注意到集合 T 是有限的. 定义 F_i 为非空有限集 $T \cap A_i$ 的凸包, 则 F_i 是 A_i 的闭的有界凸子集, 且这些集合的任何 $n+1$ 个集合构成的子类具有非空的交, 因为它至少包含 T 中的一点 (元素). 由定理 2.181, $\bigcap\limits_{i\in I} F_i$ 为非空集. 因为对所有的 i 均有 $F_i \subset A_i$, 证得结果. □

2.5.6　锥线性规划

设 X, X^* 与 Y, Y^* 是成对的局部凸的拓扑向量空间. 这一节将前一节中发展的对偶理论应用到下述形式的锥线性问题

$$(\mathrm{P})\qquad \min_{x\in C} \langle \alpha, x\rangle \quad \text{s.t.} \quad Ax + b \in K, \tag{2.345}$$

其中 $\alpha \in X^*$, $A : X \to Y$ 是连续的线性映射, $C \subset X$ 与 $K \subset Y$ 是非空的闭凸锥. 定义

$$f(x) := \begin{cases} \langle \alpha, x\rangle, & \text{若 } x \in C, \\ +\infty, & \text{若 } x \notin C. \end{cases}$$

$G(x) := Ax + b$, 上述问题成为 (2.293) 的特殊情况, 其可行集是

$$\Phi := \{x \in C : Ax + b \in K\},$$

相应的最优值函数是

$$v(y) := \inf\{\langle \alpha, x\rangle : x \in C,\ Ax + b + y \in K\}, \tag{2.346}$$

其参数化问题为

$$(\mathrm{P}_y)\qquad \min_{x\in C}\langle \alpha, x\rangle \quad \text{s.t.} \quad Ax+b+y\in K.$$

注意到, 线性问题 (2.345) 总是凸的, 因此最优值函数 $v(y)$ 亦是凸的.

下面计算 (P) 的 (Lagrange) 对偶 (见 (2.308)). 有

$$L(x,y^*)=\langle \alpha, x\rangle+\langle y^*, Ax+b\rangle$$

与

$$\inf_{x\in C} L(x,y^*)=\begin{cases}\langle y^*,b\rangle, & \text{若 } \alpha+A^*y^*\in -C^-,\\ -\infty, & \text{否则}.\end{cases}$$

于是, (2.345) 的对偶问题可以写为

$$\max_{y^*\in K^-}\langle y^*,b\rangle \quad \text{s.t.} \quad A^*y^*+\alpha\in -C^-, \tag{2.347}$$

其中 $A^*:Y^*\to X^*$ 是 A 的伴随算子. 类似地, (P_y) 的对偶是

$$(\mathrm{D}_y)\qquad \max_{y^*\in K^-}\langle y^*,b+y\rangle \quad \text{s.t.} \quad A^*y^*+\alpha\in -C^-.$$

注 2.183 问题 (D) 是另一锥线性问题, 因为我们处理的是成对的空间, 有 $A^{**}=A$, $(C^-)^-=C$, $(K^-)^-=K$, (D) 的对偶问题与 (P) 重合. 这一意义下, 问题 (P) 与 (D) 是对称的. 这一对称性基于空间 X, X^*, 与 Y, Y^* 的成对的拓扑 (paired topology). 若 X 是赋予强拓扑的自反 Banach 空间, 它可与其对偶空间是成对的, 其对偶空间 X^* 也是赋予强拓扑的 Banach 空间. 若 X 不是自反的 Banach 空间, 则 X 与 X^* 是成对的, 如果 X 赋予强拓扑, X^* 赋予弱 * 拓扑. 类似的观察也可用于空间 Y. 所以, 在非自反的 Banach 空间X(与 Y) 的情况, 它被赋予强拓扑, 我们需要赋予其对偶 X^*(与 Y^*) 弱 * 拓扑才可获得问题 (P) 与 (D) 间的对称性.

问题 (P) 是相容的, 即可行集是非空的, 当且仅当存在 $x\in C$ 满足 $Ax+b\in K$, 即 $A(C)\cap(K-b)\neq\varnothing$. 回顾, (P) 的次值是 lsc $v(0)$, 若这一次值小于 $+\infty$, 则问题 (P) 是次相容的 (subconsistent). 若空间 Y 有在 0 处的可数邻域基 (如 Y 是一赋范空间), 则 (P) 的次值可由收敛序列的极限点的下确界给出. 即 (P) 是次相容的当且仅当存在序列 $\{x_n\}\subset C$ 满足 $\lim\limits_{n\to\infty}(Ax_n+b,K)=0$, $\langle \alpha, x_n\rangle$ 是上方有界的.

根据定理 2.144, 有下述结果.

命题 2.184 *对 $y\in Y$, 成立着*

$$\mathrm{val}(D_y)=\mathrm{cl}\ v(y). \tag{2.348}$$

进一步, 若问题 (P_y) 是次相容的, 则

$$\mathrm{val}(\mathrm{D}_y) = \mathrm{lsc}\, v(y) = \min\{\mathrm{val}(\mathrm{P}_y),\ \liminf_{y' \to y} v(y')\}. \tag{2.349}$$

于是, 若问题 (P) 是次相容的, 则 (P) 与 (D) 间没有对偶间隙的充分必要条件是 (P) 的次值等于最优值 val(P). 即若 (P) 是次相容的, 则 (P) 与 (D) 间没有对偶间隙的充分必要条件是最优值函数 $v(\cdot)$ 在 $y = 0$ 处是下半连续的.

注 2.185　这里得到

$$\mathrm{val}(\mathrm{D}_y) = \sup_{y^* \in \Phi^*} \langle y^*, b + y \rangle = \sigma(b + y, \Phi^*),$$

其中

$$\Phi^* := \{y^* : y^* \in K^-,\ A^* y^* + \alpha \in -C^-\}$$

是对偶问题(D) 的可行解集. 由 (2.348) 得

$$\mathrm{cl}\, v(y) = \sigma(b + y, \Phi^*). \tag{2.350}$$

另一保证 "无对偶间隙" 的条件是 (P) 的平静性. 由命题 2.147 得, 若 (P) 是平静的, 则 (P) 与 (D) 间没有对偶间隙 , (D) 的最优解集是非空的. 根据命题 2.164 还有若 val(P) 是有限的, 或者 Y 是有限维的或者 K 的内部非空, 则 (P) 是平静的充分必要条件是

$$\liminf_{t \downarrow 0} \frac{v(td) - v(0)}{t} > -\infty, \quad \forall d \in Y. \tag{2.351}$$

考虑问题 (P_y) 的可行集

$$\Psi(y) := \{x \in C :\ Ax + b + y \in K\}. \tag{2.352}$$

注意到 $\Psi(0) = \Phi$. 若 X 和 Y 是 Banach 空间, 称 $\Psi(\cdot)$ 在 $0 \in Y$ 处是上 Lipschitz 连续的, 若

$$\Psi(y) \subset \Psi(0) + c\|y\| \overline{B}_X$$

对 $0 \in Y$ 的邻域上的所有 y 均成立, 其中 $c > 0$ 是某一常数.

命题 2.186　设 X 与 Y 是 Banach 空间, 考虑由 (2.352) 定义的多值函数$\Psi(\cdot)$. 设或者 (i) $\Psi(y)$ 在 $y = 0$ 处是上 Lipschitz 连续的; 或者 (ii) 对每一 $d \in Y$, 从 $\mathbb{R}_+$ 到 2^X 的多值函数 $t \to \Psi(td)$ 在 $t = 0$ 处是上 Lipschitz 连续的, 或者 Y 是有限维的, 或者 K 的内部是非空的. 则锥线性问题(P) 是平静的.

证明 (i) 设 $\Psi(\cdot)$ 在 $0 \in Y$ 处是上 Lipschitz 连续的. 令 $y \in Y$ 是充分接近 0 的给定点, 则或者 $\Psi(y)$ 是空集, 此种情况 $v(y) = +\infty$, 或对任意点 $x' \in \Psi(y)$, 存在 $\bar{x} \in \Phi$, 满足 $\|\bar{x} - x'\| \leqslant c\|y\|$. 结果

$$\langle \alpha, x' \rangle \geqslant \langle \alpha, \bar{x} \rangle - \|\alpha\| \, \|\bar{x} - x'\| \geqslant v(0) - c\|\alpha\| \, \|y\|.$$

因为 x' 是 $\Psi(y)$ 的任意点, 则对任何接近于 $0 \in Y$ 的 y, 可得 $v(y) \geqslant v(0) - c\|\alpha\| \, \|y\|$, 因此 $v(y)$ 在 $y = 0$ 处是次可微的, 从而 (P) 是平静的.

(ii) 考虑 $d \in Y$. 令 $t > 0$ 充分小. 则或者 $\Psi(td)$ 是空集, 从而 $v(td) = +\infty$, 或 (因为 $t \to \Psi(td)$ 在 $t = 0$ 处是上 Lipschitz 连续的) 存在正常数 $c = c(d)$ 满足对任意点 $x' \in \Psi(td)$, 存在 $\bar{x} \in \Phi$ 使得 $\|\bar{x} - x'\| \leqslant ct$ 成立. 结果

$$\langle \alpha, x' \rangle \geqslant \langle \alpha, \bar{x} \rangle - \|\alpha\| \, \|\bar{x} - x'\| \geqslant v(0) - ct\|\alpha\|,$$

从而对 $t > 0$ 充分小有 $v(td) \geqslant v(0) - ct\|\alpha\|$. 于是可推出 (2.351), 由命题 2.148 得到 (P) 的平静性.

同样的讨论可用于对偶问题(D). 即 (D) 是平静的当且仅当 val(D) 是有限的, 函数 $w(x^*)$ 在 $x^* = 0$ 处是次可微的, 其中 $w(x^*) := \inf_{y^* \in \Xi(x^*)} \langle y^*, b \rangle$,

$$\Xi(x^*) := \{y^* \in -K^- : \ A^* y^* - \alpha + x^* \in C^-\}. \tag{2.353}$$

若 (D) 是平静的, 则 (P) 与 (D) 间没有对偶间隙, (P) 的最优解集是非空的.

如注 2.183 所指出的那样, 若 X 与 Y 是 Banach 空间, X^* 与 Y^* 视为对偶 Banach 空间, 若它们分别赋予强拓扑, 除非 X 与 Y 是自反的, X 与 X^*, Y 与 Y^* 不是成对的. 因此, 对非自反的 Banach 空间, (D) 的对偶 (即双对偶问题) 与问题 (P) 是不相同的. 这说明在将命题 2.186 的结果应用于问题 (D) 的对偶时, 要十分注意.

正则性条件 (2.312), 即条件 $0 \in \operatorname{int}(\operatorname{dom} v)$ 可表示为

$$0 \in \operatorname{int}\{A(C) + b - K\}. \tag{2.354}$$

设 X 与 Y 是 Banach 空间, 由命题 2.152, 条件 (2.354) 可推出 $v(\cdot)$ 在 $y = 0$ 的连续性, 若 (P) 是次相容的, 则相反的结果由 $v(\cdot)$ 在 $y = 0$ 处的连续性可得到.

下述结果分别是定理 2.165 与命题 2.166 在此种特殊情况下的结果.

定理 2.187 设 X 与 Y 是 Banach 空间, 且正则性条件 (2.354) 成立. 则 (P) 与 (D) 间不存在对偶间隙. 进一步, 若 (P) 的最优值是有限的, 则对偶问题(D) 的最优解集 $\mathcal{S}(\mathrm{D})$ 是 Y^* 中的非空的凸的有界的弱 * 紧致子集.

命题 2.188　设 $\mathcal{S}(\mathrm{D})$ 非空, 它的回收锥是 $\{0\}$, 集合 $A(C)+b-K$ 具有非空的相对内部. 则正则性条件 (2.354) 成立, 若还有 X 与 Y 是 Banach 空间, 则 $\mathrm{val(P)}=\mathrm{val(D)}$, 且 $\mathcal{S}(\mathrm{D})$ 是有界的且弱 * 紧致的.

条件 (2.354) 对于问题 (P) 而言重合于 Robinson 约束规范. 显然, 这一条件可推出可行集 Φ 是非空的, 即问题 (P) 是相容的.

命题 2.189　设问题 (P) 是相容的. 则条件 (2.354) 等价于条件

$$A(C)-K+[\![b]\!]=Y. \tag{2.355}$$

证明　由命题 2.95, 对任意可行点 $x\in\Phi$, 条件 (2.354) 等价于

$$A(\mathcal{R}_C(x))-\mathcal{R}_K(Ax+b)=Y. \tag{2.356}$$

因为 C 与 K 是锥, 有 (见例 2.62) $\mathcal{R}_C(x)=C+[\![x]\!]$, $\mathcal{R}(Ax+b)=K+[\![Ax+b]\!]$. 所以, 条件 (2.356) 等价于条件 $A(C)-K+[\![Ax]\!]+[\![b]\!]=Y$. 观察到向量 Ax 与 $-Ax$ 属于集合 $A(C)-K+[\![b]\!]$. 事实上, 因为 x 是可行的, 有 $x\in C$, $Ax\in A(C)$. 由于 x 可行, 有 $Ax+b\in K$, 得到 $-Ax\in[\![b]\!]-K$. 由于 $A(C)-K+[\![b]\!]$ 是锥, 有 $[\![Ax]\!]\subset A(C)-K+[\![b]\!]$. 结果, 对任何可行点 x, 条件 (2.356) 等价于条件 (2.355). □

注意, 条件 (2.355) 不能推出问题 (P) 的相容性 (consistency). 例如, 取 $Y=X=C=\mathbb{R}$, $K=\mathbb{R}_+$, $b=-1$, $Ax=0$, $\forall x\in X$. 则 $Ax+b=-1$, $\forall x\in X$, 因而对应的问题 (P) 不是相容的. 另一方面, $[\![b]\!]=Y$, 因而条件 (2.355) 成立. 所以, (P) 的相容性的假设在命题 2.189 中是本质性的.

尤其, 若 C 的内部存在可行点, 则 $\mathcal{R}_C(x)=X$, 条件 (2.356) 变为

$$A(X)-K+[\![b]\!]=Y. \tag{2.357}$$

进一步, 若 $K=\{0\}$, 则这一条件简化为 $A(X)=Y$, 即 A 映 X 到 Y 上. 注意到条件 (2.357) 可由下述条件推出 (如下述例子说明, 它们非等价的)

$$A(X)-K=Y. \tag{2.358}$$

例 2.190　令 $C=X=\mathbb{R}$, $Y=\mathbb{R}^2$, $K=\{(0,y_2):y_2\geqslant 0\}$, $Ax=(x,0)\in Y$, $b=(0,1)$. 则 $x=0$ 是一可行点, 条件 (2.357) 成立, 而条件 (2.358) 不成立.

(广义)Slater 条件在此种情形成立当且仅当存在点 $\bar{x}\in C$ 满足 $A\bar{x}+b\in\mathrm{int}(K)$. 由命题 2.106 得, 若 K 的内部非空, 则条件 (2.354) 等价于 Slater 条件.

命题 2.191　设 x_0 与 $\bar{y}^*$ 分别是问题 (P) 与 (D) 的可行点. 则 x_0 是 (P) 的最优解, $\bar{y}^*$ 是 (D) 的最优解且 $\mathrm{val(P)}=\mathrm{val(D)}$ 的充分必要条件是下述互补条件成立

$$\langle\alpha+A^*\bar{y}^*,x_0\rangle=0,\quad\langle\bar{y}^*,Ax_0+b\rangle=0. \tag{2.359}$$

证明 因为 $\bar{y}^* \in N_K(Ax_0+b)$ 当且仅当 $\bar{y}^* \in K^-$ 且 (2.359) 中的第二互补条件成立. 所以, (2.359) 的第二互补条件对应着 (2.301) 中的第二最优条件. 类似地, (2.359) 中的第一互补条件对应着 (2.301) 中的第一最优条件. 由定理 2.142 即得结果. □

称问题 (P) 与 (D) 的强对偶性(strong duality) 成立, 若两个问题具有最优解且 val(P) = val(D). 上述命题表明, 强对偶性成立当且仅当 (P) 与 (D) 有满足互补条件(2.359) 的可行点.

注 2.192 若 $C = X$, 则 $C^- = \{0\}$, 则 $\bar{y}^*$ 的可行性可推出 $\alpha + A^*\bar{y}^* = 0$. 因此, 在此种情形, (2.359) 中的第一互补条件自动成立.

下述结果表明, 当 K 在 K 中某一点处的雷达锥与切锥不相等时, 可选取合适的 α, A 与 b, 使得问题的对偶间隙是存在的.

命题 2.193 *设 K 是闭凸锥, 存在一点 $\bar{y} \in K$, K 在 $\bar{y}$ 处的雷达锥与切锥不相等. 则可以构造具有形式 (2.345) 的锥线性问题(P), 满足 $C = X$ (即可找到 α, A 和 b), 使得 (P) 与对偶问题 (D) 间存在对偶间隙.*

证明 取 $X := \mathbb{R}$, $\alpha := 0$, 构造 A, b, 使得 (P) 的可行集 $\{x : Ax + b \in K\}$ 是空集, 且存在收敛到 b 的序列 $\{b_n\}$, 有 $\{x : Ax + b_n \in K\}$ 是非空的. 这种构造之下, (P) 的最优值是 $+\infty$, 而经过扰动 $y_n := b_n - b$ 的问题的最优值是 0. 因此, 此种情况, (P) 是次相容的, 定理 2.144 中的 "无对偶间隙" 条件 (2.275) 不成立. 结果, 在现在的构造之下, 对偶问题的最优值是零, 因此对偶间隙是存在的.

现在就来实现这一构造. 令 $\bar{y} \in K$ 满足 $d \in T_K(\bar{y})$ 且 $d \notin \mathcal{R}_K(\bar{y})$. 则

$$\bar{y} + td \notin K, \quad \forall t > 0. \tag{2.360}$$

取 $b := \bar{y} + b$, 定义 $A : \mathbb{R} \to Y$ 满足 $Ax := x\bar{y}$. 我们给出论断: $\forall x \in \mathbb{R}$, $Ax + b = (x+1)\bar{y} + d$ 且 $Ax + b \notin K$. 事实上, 假设这结论非真, 即存在 $\bar{x} \in \mathbb{R}$, $A\bar{x} + b \in K$. 则由于 K 是锥且 $\bar{y} \in K$ 得 $t\bar{y} + (\bar{x}+1)\bar{y} + d \in K$, $\forall t \geqslant 0$. 取 t 充分大, 则存在 $c > 0$, 使得 $c\bar{y} + d \in K$, 从而 $\bar{y} + c^{-1}d \in K$. 然而, 这与 (2.360) 是矛盾的.

为完成证明, 只需要证

$$\lim_{x \to +\infty} \operatorname{dist}(Ax + b, K) = 0. \tag{2.361}$$

由于 $d \in T_K(y)$, 对 $t \geqslant 0$ 有 $\operatorname{dist}(\bar{y} + td, K) = o(t)$. 结果,

$$\begin{aligned}\lim_{x \to +\infty} \operatorname{dist}(Ax + b, K) &= \lim_{x \to +\infty} \operatorname{dist}((x+1)\bar{y} + d, K) \\ &= \lim_{x \to +\infty} (x+1) \operatorname{dist}(\bar{y} + (x+1)^{-1}d, K) \\ &= \lim_{t \downarrow 0} t^{-1} \operatorname{dist}(\bar{y} + td, K) = 0,\end{aligned}$$

这就完成了证明. □

例 2.194 考虑下述最优化问题

$$\begin{aligned}
\min_{x,\,r}\quad & \int_0^1 s\,x(s)\mathrm{d}s + 2r, \\
\text{s.t.}\quad & 1 - \int_t^1 x(s)\mathrm{d}s - r \leqslant 0,\ t \in [0,1], \\
& r \geqslant 0,\ x(t) \geqslant 0,\ t \in [0,1],
\end{aligned} \tag{2.362}$$

其中 $r \in \mathbb{R}$, x 是在区间 $[0,1]$ 上的泛函空间 X 的元素, $(x,r) \mapsto y(t) := -\int_t^1 x(s)\mathrm{d}s - r$ 可视为从 $X \times \mathbb{R}$ 到另一泛函空间 Y 的线性映射 A. 当然, 我们需定义所考虑的 X 与 Y 是什么样的泛函空间. 取 $X = Y = L_2[0,1]$. 若定义 K 为 $L_2[0,1]$ 中的非正值函数构成的锥, $C := -K \times \mathbb{R}_+$, 则 (2.362) 具有形式 (2.345). 不难验证, $x(\cdot) \equiv 0$, $r = 1$ 是此问题的最优解, 因而其最优值是 2. 注意, 对空间 $L_2[0,1]$, 不等式 $x(t) \geqslant 0, t \in [0,1]$ 应理解为它对区间 $[0,1]$ 中的几乎处处 (以 Lebesgue 测度) 每一点处成立.

下面计算此问题的对偶. 有 $Y^* = Y$, 对 $y \in Y$, 相应的 Lagrange 函数$L(x,r,y)$ 是

$$\int_0^1 s\,x(s)\mathrm{d}s + 2r + (1-r)\int_0^1 y(t)\mathrm{d}t - \int_0^1 y(t)\left(\int_t^1 x(s)\mathrm{d}s\right)\mathrm{d}t.$$

通过交换最后一个积分的积分顺序, 得到

$$L(x,r,y) = \int_0^1 \left(s - \int_0^s y(t)\mathrm{d}t\right)x(s)\mathrm{d}s + r\left(2 - \int_0^1 y(t)\mathrm{d}t\right) + \int_0^1 y(t)\mathrm{d}t.$$

通过对 $x(\cdot) \geqslant 0$, $r \in \mathbb{R}_+$ 极小化 $L(x,r,y)$ 可导致对偶问题

$$\begin{aligned}
\max_{y}\quad & \int_0^1 y(t)\mathrm{d}t, \\
\text{s.t.}\quad & s - \int_0^s y(t)\mathrm{d}t \geqslant 0,\ s \in [0,1], \\
& \int_0^1 y(t)\mathrm{d}t \leqslant 2,\ y(s) \geqslant 0,\ s \in [0,1].
\end{aligned} \tag{2.363}$$

此对偶问题也可通过计算 A 的伴随用 (2.347) 对偶问题形式得到, 本质上需要同样的计算.

对偶问题的满足 $\int_0^1 y(t)\mathrm{d}t = 1$ 的任何可行解, 例如 $y(\cdot) \equiv 1$ 均是最优解, 对偶问题的最优解是 1. 因此, 原始问题与对偶问题间存在着对偶间隙. 为计算原始问题

(2.362) 的次值, 考虑 $L_2[0,1]$ 中的函数序列

$$x_n(t) := \begin{cases} 0, & 若\ t \in [0, 1-n^{-1}], \\ n, & 若\ t \in (1-n^{-1}, 1] \end{cases}$$

及序列 $r_n = 0$. 不难验证 $\text{dist}(A(x_n, r_n)+1, K)$ 趋于 0(距离是关于 $L_2[0,1]$ 的范数选取的) 且 $\int_0^1 s\ x_n(s)\text{d}s + 2r_n \to 1$. 因此原始问题的次值小于或等于 1. 因为对偶值是 1, 我们得到原始问题的次值实际上等于 1, 这与定理 2.144 是相吻合的.

有趣的是, 可以注意到, 若约束 $r \geqslant 0$ 从原始问题中拿掉, 则 (2.363) 中的不等式约束 $\int_0^1 y(t)\text{d}t \leqslant 2$ 被替换为等式 $\int_0^1 y(t)\text{d}t = 2$. 所以, 此种情况下, 对偶问题是不相容的, 其最优值是 $-\infty$. 可以验证原始问题相应的次值也是 $-\infty$.

现在考虑同样的原始问题 (2.362), 但取 $X = Y = C[0,1]$. 对偶空间 Y^* 是由 $[0,1]$ 上的有限符号的 Borel 测度构成的空间. 对每一测度 $\mu \in C[0,1]^*$, 定义 $F_\mu(s) := \mu([0,s])$, 由于右端是区间 $[0,1]$ 上的有界变分 (bouned variation) 函数, 这是连续函数.

原始问题还有相同的最优解和相同的最优值 2. 不难看到, 对于所选的函数空间, 原始问题满足 Slater 条件: 由类似的计算, Lagrange 函数是

$$L(x, r, \mu) = \int_0^1 (s - F_\mu(s))x(s)\text{d}s + r(2 - F_\mu(1)) + F_\mu(1).$$

因此, 对偶问题涉及的约束 $s - F_\mu(s) \geqslant 0$ 对几乎每一 $s \in [0,1]$ 成立. 因为 K^- 由非负测度 (见例 2.63) 构成, 因此, 对 $\mu \in K^-$, 函数 $F_\mu(s)$ 关于 s 是单调递增的, 这些约束等价于 $s - F_\mu(s) \geqslant 0,\ s \in [0,1]$ 结果, 对偶问题可表示为

$$\begin{aligned} \max_{\mu} \quad & F_\mu(1), \\ \text{s. t.} \quad & s - F_\mu(s) \geqslant 0,\ s \in [0,1), \\ & F_\mu(1) \leqslant 2,\ \mu \succeq 0, \end{aligned} \tag{2.364}$$

其中 $\mu \succeq 0$ 意味着测度 μ 是非负值的, 或等价地, $F_\mu(s)$ 关于 s 是单调递增的. 这一对偶问题的最优解由 $\mu = 2\delta(1)$ 给出, 其中 $\delta(1)$ 是在点 $t = 1$ 处质量 1 的测度. 所以, 对偶问题的最优值是 2, 因此原始与对偶问题间没有对偶间隙. 由于 Slater 条件, 这与定理 2.187 是一致的.

2.5.7 广义线性规划与多面多值函数

这一节用于研究线性规划, 即线性目标函数与有限多个线性等式与不等式约束的最优化问题. 这样的问题可视为当定义锥线性问题 (2.345) 中的锥 C 与 K 是多

面体的 (polyhedral) 特殊情况. 线性规划值得特别研究, 因为它们具有重要的性质且在实际应用中非常有用. 我们还考虑广义的多面集 (generalized polyhedral set) 与对应的广义线性规划问题. 下面给出多面集与广义多面集的定义.

定义 2.195 设 Y 是局部凸的拓扑向量空间. 凸集 $K \subset Y$ 被称为多面集, 若它可以表示为有限多个闭半空间的交, 即存在 $y_i^* \in Y^*$, $b_i \in \mathbb{R}$, $i = 1, \cdots, p$, 满足

$$K = \{y \in Y : \langle y_i^*, y \rangle \leqslant b_i,\ i = 1, \cdots, p\}. \tag{2.365}$$

尤其, 若 $b_i = 0$, $i = 1, \cdots, p$, 则 K 成为一凸多面锥.

凸集 $K \subset X$ 被称为广义多面集, 若它可表示为多面集与 Y 的仿射子空间的交. 即存在一闭的仿射空间 $L \subset Y$ 与 $y_i^* \in Y^*$, $b_i \in \mathbb{R}$, $i = 1, \cdots, p$, 满足

$$K = \{y \in L : \langle y_i^*, y \rangle \leqslant b_i,\ i = 1, \cdots, p\}. \tag{2.366}$$

尤其, 若 $0 \in L$, 即 L 是 Y 的闭的线性空间, $b_i = 0$, $i = 1, \cdots, p$, 则 K①成为凸的广义的多面锥.

若空间 Y 是有限维的, 则 Y 的任何仿射子空间可由有限多个线性不等式约束定义. 因此, 在有限维的情形, 广义多面集的概念与多面凸集的概念是相同的. 在无穷维空间中, 情况当然就不那么简单.

注 2.196 Y 中任何闭的仿射子空间 Y 均可以表示为 $L = \{y \in Y : Ay = z\}$, 其中 Z 是另一局部凸的拓扑向量空间, $A : Y \to Z$ 是连续的线性映射且满足 $AY = Z$, 即 A 是映上的. 为证明这一点, 考虑平行于 L 的 (闭的) 线性空间 $L' \subset Y$, 即存在 $y' \in Y$ 满足 $L = y' + L'$, 在 Y 上定义等价关系: $y_1 \sim y_2$ 当且仅当 $y_1 - y_2 \in L'$. 对应的等价类空间 $Z := Y/L'$ 赋予由 Y 引导的拓扑和代数运算成为局部凸的拓扑向量空间. 考虑映射 $A : Y \to Y/L'$, 它联系点 $y \in Y'$ 与其等价类②, 它是连续的线性映射, $L = \{y \in Y : Ay = z\}$, 其中 $z := Ay'$, A 的值域与 Y/L' 重合. 注意到, 若 Y 是 Banach 空间, L' 是 Y 中闭的线性子空间, 则空间 $Z := Y/L'$ 赋予相应的范数$\|z\| := \inf_{y \in \mathrm{equiv}(z)} \|y\|$, 其中 $\mathrm{equiv}(z)$ 记 $z \in Y/L'$ 的等价类, 也是 Banach 空间.

命题 2.197 设 $K \subset Y$ 是非空的广义多面集. 则 K 的相对内部非空.

证明 不失一般性, 设 $0 \in L$, 即仿射空间 L 是线性的. 则 L 可视为在所引导的相应的拓扑下局部凸的拓扑向量空间. 限定在空间 L 上, 广义多面集 K 成为多面集. 所以, 仅对多面集给出证明即可. 因此设 K 是 (2.365) 形式的多面集, 考虑

$$I := \{i : \exists y_i \in K,\ \langle y^*, y_i \rangle < b_i, \quad i = 1, \cdots, p\}.$$

---

① 原著中为 Y.

② 应为 $y \in Y$.

有

$$\mathrm{Sp}(K) = \{y \in Y : \langle y_i^*, y\rangle = 0,\ \forall x \notin I\},$$

因此, K 的相对内部是满足 $\langle y_i^*, y\rangle < b_i,\ \forall i \in I$ 的 $y \in K$ 的集合. 这个集合是非空的, 因为它包含点 $k^{-1}\sum_{i\in I} y_i$, 其中 $k := |I|$. □

在 (广义) 多面集上极小化线性目标的问题的两个关键性质是当最优值有限时最优解的存在性与对偶问题的 "无对偶间隙" 性质. 我们在更一般的框架下开始讨论, 证明在广义多面集上极小化凹的目标函数的最优解的存在性.

考虑下述凹的最优化问题:

$$(\mathrm{P}) \qquad \min_{x\in X} f(x) \quad \text{s. t.} \quad Ax = y_0, \quad \langle a_i, x\rangle \leqslant b_i, \quad i = 1,\cdots,p, \tag{2.367}$$

其中函数 $f : X \to \mathbb{R}$ 是凹的, $a_i \in X^*$, $b_i \in \mathbb{R}$, $i = 1,\cdots,p$, $y_0 \in Y$, $A : X \to Y$ 是从 X 到另一局部凸的拓扑向量空间 Y 上的连续的线性映射. 显然, 因为 $L := \{x : Ax = y_0\}$ 是 X 的闭的仿射子空间, 上述问题的可行集 Φ 是凸的广义多面集.

问题 (2.367) 的最优解的存在性的一个有用的刻画可通过可行集 Φ 的回收锥来描述. 注意到, 假设 Φ 是非空的, Φ 的回收锥Φ^∞ 可以表示为下述形式

$$\Phi^\infty = \{h \in X : Ah = 0,\ \langle a_i, h\rangle \leqslant 0,\ i = 1,\cdots,p\}.$$

称 h 是上述问题 (P) 在可行点 $x \in \Phi$ 处的下降方向(descent direction), 若 $h \in T_\Phi(x)$, 且对 $t > 0$ 充分小有 $f(x + th) < f(x)$.

定理 2.198 *设 $f : X \to \mathbb{R}$ 是凹函数. 则由 (2.367) 给出的问题 (P) 具有最优解的充分必要条件是最优值是有限的. 进一步, (P) 的最优值是有限的充分必要条件是下述条件成立: 问题 (P) 是相容的 (即可行集非空), 且不存在与下降方向$h \in \Phi^\infty$ 相联系的 $x \in \Phi$.*

证明 考虑下述性质: (i) (P) 的最优值有限; (ii) 问题 (P) 是相容的且不存在联系着一下降方向$h \in \Phi^\infty$ 的 x. 显然, 若 (P) 的最优解集 $\mathcal{S}(\mathrm{P})$ 非空, 则 (P) 的最优值是有限的. 因此, 只需要证 (i) $\Rightarrow$ (ii) $\Rightarrow \mathcal{S}(\mathrm{P}) \neq \varnothing$.

下证推出关系 (i) $\Rightarrow$ (ii). 只需要证若 $h \in \Phi^\infty$ 是 $x \in \Phi$ 处的下降方向, 则 $\mathrm{val}(\mathrm{P}) = -\infty$. 令 $t_0 > 0$ 满足 $f(x + t_0 h) < f(x)$. 则对任意的 $t \geqslant t_0$, 由 f 的凹性得到

$$f(x + th) \leqslant f(x) + t\Big[\frac{f(x + t_0 h) - f(x)}{t_0}\Big].$$

通过取 t 充分大, 由于 h 是 Φ 的回收方向, $x + th \in \Phi$, $t \geqslant 0$, 得到 $\mathrm{val}(\mathrm{P}) = -\infty$.

再来证推出条件 (ii) $\Rightarrow \mathcal{S}(\mathrm{P}) \neq \varnothing$. 设在 (P) 的任何可行点, 不存在回收方向 $h \in \Phi^\infty$. 设 $\{x_k\}$ 是 (P) 的极小化序列. 如果有必要, 可取一子列, 设相联系的起

作用约束集合 $I(x_k)$ 是常集合. 在满足这一性质的所有子序列中选取使 $|I_k|$ 达到最大的一个序列. 若 $\{x_k\}$ 在 $\mathcal{S}(\mathrm{P})$ 的外面, 再抽出一子序列, 可假设对所有的 $k \in \mathbb{N}$, $f(x_k) < f(x_{k-1})$. 置 $h_k := x_k - x_{k-1}$. 因为 f 是凹的, 其方向导数满足

$$f'(x_k, h_k) \leqslant -f'(x_k, -h_k) \leqslant -(f(x_{k-1}) - f(x_k)) < 0,$$

因此, 再次由 f 的凹性, 对所有的 $t > 0$, $f(x_k + th_k) < f(x_k)$. 得到 h_k 不属于 Φ^∞, 因为否则它是一下降方向, 这与问题 (P) 不具下降方向矛盾.

因为 $I(x_k) = I(x_{k-1})$, 对 $t > 0$ 充分小, $x_k + th_k \in \Phi$. 所以存在 t_k, 它是 $t = t_k$ 时有新的约束成为紧约束的最小的正数, 即 $\hat{x}_k := x_k + t_k h_k$ 满足 $I(\hat{x}_k) \supset I(x_k)$ 且 $I(\hat{x}_k) \neq I(x_k)$. 由 f 的凹性得 $f(\hat{x}_k) < f(x_k)$, 从而, 如有必要可以抽取一子序列, $\{\hat{x}_k\}$ 是满足 $I(\hat{x}_k)$ 为常集的极小化序列, 满足 $I(\hat{x}_k) \supset I(x_k)$ 且 $I(\hat{x}_k) \neq I(x_k)$, 这导致矛盾. 因为至少有一 x_k 属于 $\mathcal{S}(\mathrm{P})$, 这一集合因而是非空的. □

现在考虑下述所谓的广义线性规划(generalized linear programming) 问题

$$(\mathrm{LP}) \qquad \min_{x \in X} \langle \alpha, x\rangle \quad \text{s. t.} \quad Ax = y_0, \quad \langle a_i, x\rangle \leqslant b_i, \quad i = 1, \cdots, p. \tag{2.368}$$

取 $f(x) := \langle \alpha, x\rangle$, $\alpha \in X^*$, 则这一问题可视为凹问题 (2.367) 的特殊情况. 注意到, 如果定义 $C := X$, $K := \{0\} \times \mathbb{R}^p_- \subset Y \times \mathbb{R}^p$, 相应的仿射映射为 $x \mapsto (Ax - y_0, \langle a_1, x\rangle - b_1, \cdots, \langle a_p, x\rangle - b_p)$, 则上述问题可视为锥线性问题 (2.345). 显然, 这里的集合 K 是凸的广义多面锥.

上述问题 (LP) 的对偶可写为

$$(\mathrm{LD}) \qquad \begin{array}{ll} \max\limits_{(y^*, \lambda) \in Y^* \times \mathbb{R}^p_+} & \{-\langle y^*, y_0\rangle - \langle \lambda, b\rangle\}, \\ \text{s. t.} & \alpha + A^* y^* + \sum\limits_{i=1}^n \lambda_i a_i = 0. \end{array} \tag{2.369}$$

注意到这一对偶问题是锥线性问题 (2.345) 在目前的特殊情况下 (LP) 的对偶问题 (2.347) 的具体形式. 这是因为 $C := X$ 的极锥是 $\{0\}$, $K := \{0\} \times \mathbb{R}^p_-$ 的极锥是 $Y^* \times \mathbb{R}^p_+$.

由于任何线性函数均是凹函数, 可将定理 2.198 应用到广义线性规划问题 (2.368), 得到下述结果.

定理 2.199 *设由 (2.368) 给出的广义线性规划问题 (LP) 是相容的. 则下述结论成立: 若存在 $h \in X$, 满足*

$$\langle \alpha, h\rangle < 0, \quad Ah = 0 \text{ 且 } \langle a_i, h\rangle \leqslant 0, \quad i = 1, \cdots, p, \tag{2.370}$$

则 $\mathrm{val(P)} = -\infty$; 否则, $\mathrm{val(P)}$ 是有限的, (LP) 具有非空的最优解集.

上述定理表明, 广义线性规划问题有最优解的充分必要条件是它的最优值是有限的.

下述定理被经常称为 Hoffman 引理. 它的意思是广义线性规划问题的可行解集关于右端参数的扰动为 Lipschitz 连续的.

定理 2.200 (Hoffman 引理) 设 X 与 Y 是 Banach 空间, 令 $A: X \to Y$ 是线性的连续映射, 其值域是闭的(即值域 AX 是 Y 的闭子空间). 给定 $x_i^* \in X^*$, $i = 1, \cdots, p$. 考虑多值函数

$$\Psi(y, b) := \{x \in X : Ax = y, \ \langle x_i^*, x\rangle \leqslant b_i, \ i = 1, \cdots, p\}. \tag{2.371}$$

则存在依赖于 A 与 x_i^*, $i = 1, \cdots, p$ 的常数 $k > 0$, 满足对任意 $x \in X$, $(y, b) \in \operatorname{dom} \Psi$,

$$\operatorname{dist}(x, \Psi(y, b)) \leqslant k\left(\|Ax - y\| + \sum_{i=1}^{p}[\langle x_i^*, x\rangle - b_i]_+\right). \tag{2.372}$$

证明 首先, 不失一般性, 可设 $A: X \to Y$ 是映上的. 实际上, 因为 $Z := AX$ 是闭的, 可视它为赋予相应的引导范数的 Banach 空间. 显然, 若 $y \notin Z$, $\Psi(y, b)$ 是空集. 因此, 在不等式 (2.372) 的证明中, 可将 Y 代替为 Z.

其次, 只需给出当 y 固定时的证明. 即设对给定的 y, 不等式 (2.372) 对 k 取依赖于 x_i^*, $i = 1, \cdots, p$ 的 κ 值成立. 令 $b = (b_1, \cdots, b_p)$ 满足 $(y, b) \in \operatorname{dom} \Psi$, 即 $\Psi(y, b) \neq \varnothing$. 因为 A 是映上的, 由于开映射定理2.70, 存在 $\varepsilon > 0$ 满足 $\varepsilon B_Y \subset AB_X$. 因此, 给定任意 $x \in X$, 存在 $x_1 \in X$ 满足 $Ax_1 = y - Ax$ 且 $\|x_1\| \leqslant \varepsilon^{-1}\|y - Ax\|$, 因而 $x_2 := x + x_1$ 满足 $Ax_2 = y$ 且

$$\begin{aligned}\operatorname{dist}(x, \Psi(y, b)) &\leqslant \operatorname{dist}(x_2, \Psi(y, b)) + \|x - x_2\| \\ &\leqslant \operatorname{dist}(x_2, \Psi(y, b)) + \varepsilon^{-1}\|Ax - y\|.\end{aligned}$$

因为 $Ax_2 = y$, 由上述假设可得

$$\operatorname{dist}(x_2, \Psi(y, b)) \leqslant \kappa\left(\sum_{i=1}^{p}[\langle x_i^*, x_2\rangle - b_i]_+\right).$$

进一步有

$$\sum_{i=1}^{p}[\langle x_i^*, x_2\rangle - b_i]_+ \leqslant \|x - x_2\|\sum_{i=1}^{p}\|x_i^*\| + \sum_{i=1}^{p}[\langle x_i^*, x\rangle - b_i]_+,$$

且 $\|x - x_2\| \leqslant \varepsilon^{-1}\|Ax - y\|$. 结合上述不等式, 得到要证的结果, 其中 k 定义为

$$k := \kappa\left(1 + \varepsilon^{-1}\sum_{i=1}^{p}\|x_i^*\|\right) + \varepsilon^{-1}$$ ①

① 原著中 p 为 n.

剩下要证, 对于 y 是固定的, 当 $Ax = y$, 不等式 (2.372) 成立. 不失一般性, 可假设 $y = 0$. 通过限定空间 X 到 $\ker A$, 可以从证明中把映射 A 移去. 因此, 在后续证明中, 多值函数 Ψ 仅由线性不等式约束定义. 当然, 此种情形, 多值函数 Ψ 仅依赖于 $b \in \mathbb{R}^p$. 考虑映射 $M(\cdot) := (\langle x_1^*, \cdot\rangle, \cdots, \langle x_p^*, \cdot\rangle) : X \to \mathbb{R}^p$. 因为 M 的值域是有限维的, 空间 X 具有有限维子空间 X' 满足 $X = (\ker M) \oplus X'$, 即 $(\ker M) \cap X' = \{0\}$ 且 $X = (\ker M) + X'$. 则对任意 $x \in X$, 存在 (唯一的)$x_1 \in \ker M$, $x_2 \in X'$ 满足 $x = x_1 + x_2$, 有 $\operatorname{dist}(x, \Psi(b)) \leqslant \operatorname{dist}(x_2, \Psi(b) \cap X')$. 因此, 只需对 $x \in X'$ 给出证明. 可设空间 X 是有限维的. 进一步. 由于 X 是有限维的, 可赋予它 Euclid 范数, 这一范数与原范数是等价的, 因此视 X 与其对偶 X^* 是相同的.

最后, 考虑点 $x \in X$, $b \in \mathbb{R}^p$ 满足 $\Psi(b)$ 非空. 令 x' 是 x 到集合 $S := \Psi(b)$ 上的直交投影, 即 x' 是 S 中的离 x (以 Euclid 范数) 最近的点. 则对任意 $h \in \mathcal{R}_S(x')$, 有

$$0 \leqslant \lim_{t \downarrow 0} \frac{\|x' + th - x\|^2 - \|x' - x\|^2}{2t} = \langle h, x' - x\rangle.$$

因此, $(x - x') \in N_S(x')$. 由于引理 2.42, 存在满足 $\lambda_i(\langle x_i^*, x\rangle - b_i) = 0$, $\forall i = 1, \cdots, p$ 的 $\lambda_i \in \mathbb{R}_+$, 使得 $x - x' = \sum_{i=1}^p \lambda_i x_i^*$. 于是

$$\|x - x'\|^2 = \sum_{i=1}^p \lambda_i \langle x_i^*, x - x'\rangle = \sum_{i=1}^p \lambda_i(\langle x_i^*, x\rangle - b_i) \leqslant \|\lambda\|_\infty \sum_{i=1}^p [\langle x_i^*, x\rangle - b_i]_+.$$

由命题 2.41, 存在不依赖于 x 的常数 $c > 0$, $\|\lambda\|_\infty \leqslant c\|x - x'\|$. 结果

$$\|x - x'\| \leqslant c \sum_{i=1}^p [\langle x_i^*, x\rangle - b_i]_+,$$

这完成证明. □

注意到, 由 (2.371) 定义的集合 $\Psi(y, b)$ 也可写为下述形式

$$\Psi(y, b) = \{x \in X,\ G(x, y, b) \in K\}, \tag{2.373}$$

其中 $G(x, y, b) := (Ax - y, \langle x_1^*, x\rangle - b_1, \cdots, \langle x_p^*, x\rangle - b_p)$, $K := \{0\} \times \mathbb{R}_-^p$. 则对 $(y, b) \in \operatorname{dom}(\Psi)$, 若在 $\mathbb{R}^p$ 中用 l_1 范数, 不等式 (2.372) 具有下述形式

$$\operatorname{dist}(x, \Psi(y, b)) \leqslant k\ \operatorname{dist}(G(x, y, b), K). \tag{2.374}$$

这种形式下, 上界 (2.374) 类似于度量正则性(见定理 2.83) 的上界. 然而, 也有两个本质性的区别. 首先, 上界 (2.374) 是全局性的, 即对所有的 $x \in X$, $(y, b) \in \operatorname{dom} \Psi$,

这一不等式是成立的. 第二, 它不需要正则性条件, 而且对 (y,b) 的某些值 $\Psi(y,b)$ 可能是空集. 另一方面, 这里的映射 G 具有特殊 (线性) 结构, 而且 K 是广义多面集.

命题 2.201 (广义 Farkas 引理) 设 X 与 Y 是 Banach 空间, $A: X \to Y$ 是具有闭值域的线性的连续映射, $a_i \in X^*$, $i = 1, \cdots, p$. 则下述成立:

(i) 集合 (锥)

$$K := \{x \in X:\ Ax = 0,\ \langle a_i, x\rangle \leqslant 0,\ i = 1, \cdots, p\}$$

的极锥可写为

$$K^- = A^*(Y^*) + \left\{\sum_{i=1}^{p} \lambda_i a_i:\ \lambda_i \geqslant 0,\ i = 1, \cdots, p\right\}. \tag{2.375}$$

(ii) 存在 $c > 0$ 满足对每一 $x^* \in K^-$, 存在 $y^* \in Y^*$, $\lambda \in \mathbb{R}^p_+$,

$$\|y^*\| + \|\lambda\| \leqslant c\|x^*\| \text{ 且 } x^* = A^*y^* + \sum_{i=1}^{p} \lambda_i a_i \text{①} \tag{2.376}$$

证明 (i) 令 $x^* \in K^-$. 考虑在 K 上极小化 $-\langle x^*, \cdot\rangle$ 的线性的锥问题. 这一问题的对偶是在 (2.375) 右端的集合上极大化常数 0. 结合 Hoffman 引理(定理 2.200) 与命题 2.186, 得到原始问题是平静的; 因此对偶问题的可行集是非空的, 或等价地, K^- 被包含在 (2.375) 的右端. 相反的包含关系显然也是成立的, 因此 (2.375) 成立.

(ii) 令 $\hat{a}_i$, $i = 1, \cdots, p$ 是 a_i 到 $\ker A$ 上的限制. 令 $x^* := A^*y^* + \sum_{i=1}^{p} \lambda_i a_i$, $\lambda \in \mathbb{R}^p_+$. 记 $\hat{x}^*$ 为 x^* 到 $\ker A$ 上的限制. 因为由引理 2.173 有 $(\ker A)^\perp = \operatorname{range}(A^*)$, 得 $\hat{x}^* = \sum_{i=1}^{p} \lambda_i \hat{a}_*$. 由命题 2.41, 存在不依赖于 x^* 的 $c_1 > 0$ 与 $\hat{\lambda} \in \mathbb{R}^p_+$②满足 $\hat{x}^* = \sum_{i=1}^{p} \hat{\lambda}_i \hat{a}_i$ 及 $\|\hat{\lambda}\| \leqslant c_1\|\hat{x}^*\|$. 由 Hahn-Banach 定理, 存在 $\hat{x}^*$ 的延拓 $\tilde{x}^* \in X^*$ 满足 $\|\tilde{x}^*\| = \|\hat{x}^*\| \leqslant \|x^*\|$. 由于对所有的 $x \in \ker A$ 有 $\langle \tilde{x}^*, x\rangle = \langle \hat{x}^*, x\rangle = \langle x^*, x\rangle$, 有 $x^* - \tilde{x}^* \in (\ker A)^\perp = \operatorname{range}(A^*)$. 由引理 2.173, 存在 $\hat{y}^* \in Y^*$ 满足 $x^* - \tilde{x}^* = A^*\hat{y}^*$, 存在不依赖于具体点的 $c_2 > 0$ 满足 $\|\hat{y}^*\| \leqslant c_2\|x^* - \tilde{x}^*\| \leqslant 2c_2\|x^*\|$. 令 $c := c_1 + 2c_2$, 即证得 (2.376). □

定理 2.202 (广义线性规划的对偶定理) 设 X 与 Y 是 Banach 空间, 令 $A: X \to Y$ 是值域为闭的线性的连续映射. 设问题 (LP) 是相容的. 则 (LP) 与 (LD) 的最优值是相等的, 若其公共最优值是有限的, 则这两个问题均有非空的最优解集.

---

① 原著中 p 为 n.

② 原著中为 $c > 0$.

证明　因为 (LP) 是相容的, $\mathrm{val(LP)} < +\infty$. 因为 $\mathrm{val(LD)} \leqslant \mathrm{val(LP)}$, 则若 $\mathrm{val(LP)} = -\infty$, 结论当然成立. 否则, (LP) 有有限的最优值, 由定理 2.198 得, 它有最优解. 结合 Hoffman 引理 (定理 2.200) 与命题 2.186, 得到 (LP) 是平静的. 结果, (LP) 与 (LD) 具有相同的最优值, $\mathcal{S}(\mathrm{LD})$ 是非空的. □

下述例子表明, 即使在有限维的情况, 也可能发生 (LP) 与 (LD) 非相容的情形 (即有空的可行集), 此种情形, 当然有 $-\infty = \mathrm{val(LD)} \neq \mathrm{val(LP)} = +\infty$.

例 2.203　考虑线性规划问题

$$\min_{x \in \mathbb{R}} -x \quad \text{s. t.} \quad 0 \times x = -1, \quad x \geqslant 0.$$

这一问题是不相容的, 其对偶问题

$$\max_{\lambda,\ s \in \mathbb{R}} \lambda \quad \text{s. t.} \quad -1 + 0 \times \lambda = s, \quad s \geqslant 0$$

也不相容。

定理 2.204 (广义线性规划的第二对偶定理)　*设 X 与 Y 是 Banach 空间, $A : X \to Y$ 是具有闭值域的线性的连续映射. 设或者空间 X 是自反的或者 Y 是有限维的. 则除非两个问题是不相容的, 即 $\mathrm{val(LP)} = +\infty$ 且 $\mathrm{val(LD)} = -\infty$, (LP) 与 (LD) 的最优值彼此相等.*

证明　根据定理 2.202, 只需证若 val(LD) 是有限的, 则 (LP) 是相容的. 不难由命题 2.201(ii) 得映射 $(y^*, \lambda) \mapsto A^* y^* + \sum_{i=1}^{p} \lambda_i a_i$ 具有闭的值域. 视问题 (LD) 为原始问题, 可用定理 2.202. 在双对偶空间里, (LD) 的对偶问题可表示为

$$\text{(LDD)} \qquad \min_{x^{**} \in X^{**}} \langle x^{**}, \alpha \rangle \quad \text{s.t.} \quad A^{**} x^{**} = y_0, \quad \langle x^{**}, a_i \rangle \leqslant b_i, \quad i = 1, \cdots, p,$$

其中 $A^{**} : X^{**} \to Y^{**}$ 是 A 的双伴随算子 (biadjoint)(注意到 Y 可以自然地嵌入到 Y^{**}). 若 X 是自反空间, 则 (LDD) 与 (LP) 重合, 因而 $\mathrm{val(LD)} = \mathrm{val(P)}$, 这完成了此种情形的证明.

现在考虑 Y 是有限维的情况. 因为 (LDD) 有有限的最优值, 它具有最优解 $x^{**} \in X^{**}$. 由引理 2.22 得, 存在 $x \in X$ 满足 $\langle \alpha, x \rangle = \langle x^{**}, \alpha \rangle$, $Ax = A^{**} x^{**}$ 且 $\langle a_i, x \rangle = \langle x^{**}, a_i \rangle$, $i = 1, \cdots, p$. 可推出 $\mathrm{val(LP)} \leqslant \mathrm{val(LDD)}$, 因此 (LP) 是相容的, 完成证明. □

例 2.205　考虑问题

$$\text{(LP}')\qquad \min_{x \in X} \langle \alpha, x \rangle \quad \text{s. t.} \quad A_1 x + y_0 \in K, \tag{2.377}$$

其中 X, Y, Z 是 Banach 空间, $A_1 : X \to Y$, $A_2 : Y \to Z$ 是连续的线性映射,

$$K := \{y \in Y : \ A_2 y = z_0, \ \langle y_i^*, y\rangle \leqslant b_i, \ i = 1, \cdots, p\},$$

即 K 是空间 Y 的一广义多面集. (LP$'$) 的可行集可表示为下述等价形式

$$\{x \in X : \ A_2 \circ A_1 x = z_0 - A_2 y_0, \ \langle A_1^* y_i^*, x\rangle \leqslant b_i - \langle y_i^*, y_0\rangle, \ i = 1, \cdots, p\},$$

其中 $A_2 \circ A_1 : \ X \to Z$ 是相应的复合映射.

从而 (LP$'$) 也是广义的线性规划问题, 其相应的约束由上面给出. 若 $A_2 \circ A_1$ 具有闭的值域, 则由定理 2.202 得, 存在 (LP$'$) 的最优解 $\bar{x}$, 至少联系着对偶最优解 $(z^*, \lambda) \in Z^* \times \mathbb{R}_+^p$(被称为 Lagrange 乘子, 见 3.1.1 节), 满足

$$\alpha + A_1^* \circ A_2^* z^* + \sum_{i=1}^{p} \lambda_i A_1^* y_i^* = 0, \tag{2.378}$$

$$\lambda_i(\langle A_1^* y_i^*, \bar{x}\rangle - b_i + \langle y_i^*, y_0\rangle) = 0, \quad i = 1, \cdots, p. \tag{2.379}$$

我们可以把方程 (2.378) 表示为 $\alpha + A_1^* y^* = 0$, 其中 $y^* := A_2^* z^* + \sum_{i=1}^{p} \lambda_i y_i^* \in Y^*$. 考虑互补条件(2.379), 有 $y^* \in N_K(A_1 x + y_0)$, 这意味着 (y^*, λ) 是原始形式 (2.377) 给出的约束的 Lagrange 乘子. 已经证得, 若 $A_2 \circ A_1$ 具有闭的值域且 (LP$'$) 的最优值是有限的, 则其对偶问题至少有一最优解.

现在讨论 Hoffman 引理的另一结论.

定义 2.206 *多值函数 $F : X \to 2^Y$ 称为 (广义的) 多面集的, 若其图是 $X \times Y$ 中有限多个被称为组分 (components) 的 (广义) 多面集的并. 在只有一个组分的情况, 即 $\mathrm{gph}(F)$ 是 (广义) 的凸多面集, 称 F 是 (广义) 凸多面集的.*

考虑广义的凸多面集的多值函数$F : X \to 2^Y$. 由上述定义, 它的图可表示为如下形式

$$\mathrm{gph}(F) = \{(x, y) \in L : \langle x_i^*, x\rangle + \langle y_i^*, y\rangle \leqslant b_i, \ i = 1, \cdots, p\}, \tag{2.380}$$

其中 L 是 $X \times Y$ 的仿射子空间, $x_i^* \in X^*$, $y_i^* \in Y^*$, $b_i \in \mathbb{R}$, $i = 1, \cdots, p$. 由注 2.196 仿射空间 L 可表示为下述形式

$$L := \{(x, y) \in X \times Y : \ A_1 x + A_2 y = z\}, \tag{2.381}$$

其中 (A_1, A_2) 是从 $X \times Y$ 映到另一 Banach 空间 Z 的连续的线性映射. 下述结果表明, 假设映射 $A_2 : Y \to Z$ 具有闭的值域, 多值函数 F 以 Hausdorff 度量 (在其定义域上) 是 Lipschitz 连续的. 当然, A_2 具有闭的值域这一技术性条件在多值函数 F 是凸多面集的情况下是不需要的.

注意到, 在注 2.196 的构造中, $Z := (X \times Y)/L'$, 其中 L' 是平行于 L 的线性空间, A_2 的值域由类 $(0, y) + L'$, $y \in Y$ 的集给出. 因此, 若空间 $\{0\} \times Y + L$ 是闭的, 此种情况下, A_2 有闭的值域.

定理 2.207　*设 X 与 Y 是 Banach 空间, $F : X \to 2^Y$ 是广义的凸的多面集的多值函数, 其图由 (2.380) 给出, 仿射空间 L 表示为 (2.381) 形式. 设映射 A_2 具有闭的值域. 则 F 的定义域是 X 的广义凸的多面集, 且存在常数 $c > 0$ 满足对任意 x_1, $x_2 \in \operatorname{dom}(F)$, 有*

$$\operatorname{Haus}(F(x_1), F(x_2)) \leqslant c\|x_1 - x_2\|. \tag{2.382}$$

证明　令 x_1, $x_2 \in \operatorname{dom}(F)$, 考虑一点 $\bar{y} \in F(x_2)$. 有

$$F(x_1) = \{y :\ A_2 y = z - A_1 x_1,\ \langle y_i^*, y\rangle \leqslant b_i - \langle x_i^*, x_1\rangle,\ i = 1, \cdots, p\},$$

因此由 Hoffman 定理 (定理 2.200) 得, 存在常数 k, 仅依赖于 A_2, $y_1^*, \cdots, y_p^*$, 满足

$$\operatorname{dist}(\bar{y}, F(x_1)) \leqslant k\left(\|A_1 x_1 + A_2 \bar{y} - z\| + \sum_{i=1}^{p} [\langle y_i^*, \bar{y}\rangle + \langle x_i^*, x_1\rangle - b_i]_+\right).$$

由于 $\bar{y} \in F(x_2)$, 有

$$\|A_1 x_1 + A_2 \bar{y} - z\| = \|A_1(x_1 - x_2)\|,$$

$$[\langle y_i^*, \bar{y}\rangle + \langle x_i^*, x_1\rangle - b_i]_+ \leqslant \langle x_i^*, x_1 - x_2\rangle_+, \quad i = 1, \cdots, p.$$

结果

$$\operatorname{dist}(\bar{y}, F(x_1)) \leqslant k\left(\|A_1(x_1 - x_2)\| + \sum_{i=1}^{p} \langle x_i^*, x_1 - x_2\rangle_+\right) \leqslant c\|x_1 - x_2\|,$$

其中 $c := k\left(\|A_1\| + \sum_{i=1}^{p} \|x_i^*\|\right)$ 不依赖于 x_1, x_2 或 $\bar{y} \in F(x_2)$. 则不等式 (2.382) 成立.

剩下需要证集合 $\operatorname{dom}(F)$ 是广义的凸的多面集. 考虑函数

$$f(x) := \inf_{y \in Y} \left\{ \max(\|A_1 x + A_2 y - z\|, \max_{1 \leqslant i \leqslant p} \{\langle x_i^*, x\rangle + \langle y_i^*, y\rangle - b_i\}) \right\}.$$

需验证 $\operatorname{dom}(F) = \{x \in X : f(x) \leqslant 0\}$. 显然, 若 $x \in \operatorname{dom}(F)$, 则 $f(x) \leqslant 0$. 相反地, 设 $f(x) \leqslant 0$. 则存在 Y 中的序列 $\{y_n\}$ 满足 $\|A_1 x + A_2 y_n - z\| \to 0$ 及 $\langle x_i^*, x\rangle + \langle y_i^*, y_n\rangle - b_i \leqslant o(1)$. 这意味着 $A_1 x - z$ 属于 $\operatorname{range}(A_2)$ 的闭包. 由于后

者是闭的, 有 $A_1x - z \in \text{range}(A_2)$, 因此存在序列 $\hat{y}_n \in Y$, 满足 $\hat{y}_n = y_n + o(1)$ 与 $A_1x + A_2\hat{y}_n - z = 0$. 所以, 广义线性规划问题

$$\min_{(\alpha,y)\in\mathbb{R}\times Y} \alpha \quad \text{s. t.} \quad A_1x + A_2y - z = 0,$$
$$\langle x_i^*, x\rangle + \langle y_i^*, y\rangle \leqslant b_i + \alpha, \quad i = 1, \cdots, p$$

具有非正值. 由定理 2.198, 只要最优值有限, 这一问题就有最优解, 因而存在可行点 (α, y) 满足 $\alpha \leqslant 0$, 这意味着 $y \in F(x)$. 这证得 $\text{dom}(F) = \{x \in X : f(x) \leqslant 0\}$.

函数 f 的一个等价表示为

$$f(x) = \inf_{y\in Y} \max_{(\lambda,\gamma)\in E} \left\{ \langle \lambda, A_1x + A_2y - z\rangle + \sum_{i=1}^p \gamma_i(\langle x_i^*, x\rangle + \langle y_i^*, y\rangle - b_i) \right\},$$

其中 $E := \left\{ (\lambda, \gamma) \in Z^* \times \mathbb{R}_+^p : \ \|\lambda\| + \sum_{i=1}^p \gamma_i = 1 \right\}$. 由 Banach-Alaoglu 定理2.27, 作为紧致集合乘积的闭子集, 集合 E 是紧致的. 因此, 由命题 2.176,

$$\begin{aligned} f(x) &= \max_{(\lambda,\gamma)\in E} \inf_{y\in Y} \left\{ \langle \lambda, A_1x + A_2y - z\rangle + \sum_{i=1}^p \gamma_i(\langle x_i^*, x\rangle + \langle y_i^*, y\rangle - b_i) \right\} \\ &= \max_{(\lambda,\gamma)\in E'} \left\{ \langle \lambda, A_1x - z\rangle + \sum_{i=1}^p \gamma_i(\langle x_i^*, x\rangle - b_i) \right\}, \end{aligned}$$

其中 $E' := \left\{ (\lambda, \gamma) \in E : \ A_2^*\lambda + \sum_{i=1}^p \gamma_i y_i^* = 0 \right\}$. 若 E' 是空集, 则 $f(x) \equiv -\infty$, 即 $\text{dom}(F) = X$. 否则, 由命题 2.3(ii), E' 是紧致集, 因此 f 是连续的. 令 $E'' := \mathbb{R}_+(E')$. 有 $\text{dom}(F)$ 与下述集合重合

$$\left\{ x \in X : \ \langle \lambda, Ax_1 - z\rangle + \sum_{i=1}^p \gamma_i(\langle x_i^*, x\rangle - b_i) \leqslant 0, \ \forall(\lambda, \gamma) \in E'' \right\}.$$

令 $z^* \in \ker(A_2^*)$. 则如果 $(\lambda, \gamma) \in E''$, $\forall t \in \mathbb{R}$, $(\lambda + tz^*, \gamma) \in E''$, 因此若 $x \in \text{dom}(F)$, 则 $\langle z^*, A_1x - z\rangle = 0$. 此等式对 $\ker(A_2^*)$ 的任意一元素均成立的充分必要条件是 $A_1x - z \in \ker(A_2^*)^\perp$. 根据引理 2.173(iv), $\ker(A_2^*)^\perp = \text{range}(A_2)^\perp$, 有 $\text{dom}(F) \subset X_1$, 其中 $X_1 := \{x \in X : A_1x - z \ \in \text{range}(A_2)\}$. 实际上, 这由 F 的定义是显然的, 但更重要的是, 对闭的仿射空间 X_1 中的 x, 满足 $(\lambda, \gamma) \in E''$ 的 λ 的选取并不重要. 考虑集合

$$\Gamma := \left\{ \gamma \in \mathbb{R}_+^p : \sum_{i=1}^p \gamma_i = 1, \ \sum_i \gamma_i y_i^* \in \text{range}(A_2^*) \right\}.$$

因为 $\mathrm{range}(A_2^*) = \ker(A_2)^\perp$, 集合 Γ 是多面集且有界, 因此是由其极点的凸组合组成的, 将 (有限的) 极点集记为 Γ^*. 对每一 $\gamma \in \Gamma^*$, 我们联系一乘子 $\lambda(\gamma)$ 使得 $(\lambda(\gamma), \gamma) \in E''$. 则 $\mathrm{dom}(F)$ 等于

$$X_1 \cap \left\{ x \in X : \langle \lambda(\gamma), A_1 x - z \rangle + \sum_{i=1}^{p} \gamma_i (\langle x_i^*, x \rangle - b_i) \leqslant 0,\ \forall \gamma \in \Gamma^* \right\}.$$

因为 Γ^* 是有限集, 由上述关系显然有集合 $\mathrm{dom}(F)$ 是广义多面集, 这证得结论. □

对非凸的多面集的多值函数, Lipschitz 连续性 (2.382) 不一定成立. 例如, 考虑下述从 $\mathbb{R}$ 到 $2^{\mathbb{R}}$ 的多值函数. 定义任给 $x \neq 0$, $F(x) := \{0\}$, $F(0) := \mathbb{R}$. 得到 $\mathrm{gph}(F) = (\mathbb{R} \times \{0\}) \cup (\{0\} \times \mathbb{R})$, F 是具有两个组分的多面集的映射. 很显然, F 在零的任何邻域内均不是 (以 Hausdorff 度量) Lipschitz 连续的. 原因很清楚: 当 x 离开 0 时, 对应图的组分 $\{0\} \times \mathbb{R}$ 的 $F(x)$ 的组分成为空集. 然而, 此例的多值函数 F 在每一点处均是上 Lipschitz 连续的. 上 Lipschitz 连续性这一性质对任何多面集的多值函数 F 均是成立的. 这通过应用定理 2.207 于 F 的每一组分上, 并注意到 F 的定义域是闭的, 可立即得到. 相应的常数取为相应 F 的所有组分的 Lipschitz 常数的最大值, 因此对所有的 x 均是相同的. 于是得到下述结果.

定理 2.208 *设 X 与 Y 是 Banach 空间, $F : X \to 2^Y$ 是多面集的多值函数. 则存在常数 $c > 0$, 满足: F 在每一点 $x \in X$ 处均是模 c 的上 Lipschitz 连续的.*

例 2.209 设 K 是 Banach 空间 Y 中具有 (2.366) 形式的 (广义的) 凸多面集. 则法锥多值函数 $y \mapsto N_K(y)$ 是 (广义的) 多面集的 (不必是凸的). 事实上, 定义 K 的面

$$K_I := \{y \in Y :\ \langle y_i^*, y \rangle = b_i,\ \forall i \in I\}, \tag{2.383}$$

其中 I 是 $\{1, \cdots, p\}$ 的子集. 显然 K 是其非空面的并. 每一面均是 (广义的) 凸多面集, 因此有非空的相对内部. 事实上, K 是其非空面的相对内部的并集. 令 $J := \{1 \leqslant j \leqslant p; 对\ \forall y \in K_I,\ \langle y_j^*, y \rangle = b_i\}$. 则 $I \subset J$ 在一面 K_I 的相对内部上的法锥是常值的, 等于

$$N_I := \mathrm{range}(A^*) + \left\{ \sum_{i \in J} \lambda_i y_i^*,\ \lambda_i \in \mathbb{R}_+^{|I|} \right\}.$$

因为若 y 在 K_I 的相对边界上, $N_I \subset N_K(y)$, 得到 $\mathrm{gph}(N_K) = \bigcup_{I \subset \{1, \cdots, p\}} K_I \times N_I$, 结果得证.

例 2.210 设 X 是 Banach 空间, $f : X \to \overline{\mathbb{R}}$ 是正常的 (广义) 凸的多面值函数, 即 f 是正常的, 其上图是非空的 (广义) 凸的多面集. 则多值函数 $x \mapsto \partial f(x)$ 是

(广义的) 多面集的 (不必是凸的). 实际上, 由上面的结论, f 的上图的法锥是 (广义的) 多面集的多值函数. 另一方面, 由注 2.117,

$$\operatorname{gph}(\partial f) \times \{-1\} = \operatorname{gph}(N_{\operatorname{epi}(f)}) \cap (X \times \mathbb{R} \times X^* \times \{-1\}),$$

由此得到结论.

例 2.211 考虑由 (2.368) 定义的广义线性规划(LP_y) 与其对偶 (LD_y), 上述记号中, 我们强调对向量 $y = y_0$ 的依赖性. 设 X 与 Y 是 Banach 空间, A 有闭的值域. 将 $v(y)$ 记以 y 为参数的最优值函数. 则在其定义域上, $v(y)$ 等于对偶问题的最优值. 对偶问题的可行集是不依赖于 y 的, $\mathcal{S}(\mathrm{LD}_y)$ 是对偶可行集的一个面. 这样的面的个数是有限的, 可以如下计算最优值: 把面从 1 到 q 排号, 选取对偶可行点 (y_k^*, λ_k), $k = 1, \cdots, q$, 满足 (y_k^*, λ_k) 属于第 k 面的相对内部. 则可以得到

$$v(y) = I_{\operatorname{dom}(v)}(y) + \max_k\{-\langle y^*, y_k\rangle - \langle \lambda_k, b\rangle\}.$$

由定理 2.207, $\operatorname{dom}(v)$ 是广义的多面集, 显然 v 是一广义的凸的多面集函数. 因为 $\partial v(y) = \mathcal{S}(\mathrm{LD}_y)$, 由例 2.210 中的讨论, 映射 $y \mapsto \mathcal{S}(\mathrm{LD}_y)$ 是广义的多面集的多值函数.

尤其, 我们可将 v 的定义域表示为有限多个广义多面集 $Y_1, \cdots, Y_q$ 的并, 且存在 Lagrange 乘子 (y_k^*, λ_k), $k = 1, \cdots, q$ 满足对每一 $y \in Y_k$, (y_k^*, λ_k) 是对偶最优解. 由互补条件, 对 $y \in Y_k$, $\mathcal{S}(\mathrm{LP}_y)$ 可表示为

$$\{x \in X : Ax = y,\ \langle a_i, x\rangle \leqslant b_i,\ \lambda_k(\langle a_i, x\rangle - b_i) = 0,\ i = 1, \cdots, p\}.$$

由于这一表示定义了广义的多面凸的多值映射, 此多值函数的图是这样的集合的并, 得到 $y \mapsto \mathcal{S}(\mathrm{LP}_y)$ 是广义的多面集的多值函数.

第3章　最优性条件

这一章讨论下述最优化问题的一阶与二阶最优性条件

$$\text{(P)} \qquad \min_{x \in Q} f(x) \qquad \text{s. t.} \qquad G(x) \in K, \tag{3.1}$$

其中 $f: X \to \mathbb{R}$, $G: X \to Y$, Q 与 K 是 X 与 Y 的非空的闭的凸子集. 可视集合 Q 为目标函数 f 的定义域. 除非特别说明, 设 X 与 Y 是 Banach 空间, $f(x)$ 与 $G(x)$ 是连续的. 记

$$\Phi := \{x \in Q: \ G(x) \in K\} = Q \cap G^{-1}(K) \tag{3.2}$$

与

$$L(x, \lambda) := f(x) + \langle \lambda, G(x) \rangle, \ (x, \lambda) \in X \times Y^* \tag{3.3}$$

分别是 (P) 的可行集与 Lagrange 函数.

首先研究一阶, 尤其是定义是下述意义的二阶增长条件(second order growth). 回顾集合 $S \subset X$ 的邻域是 X 的子集, 它是 S 的每一点的邻域. 这意味着 S 被包含在其邻域的内部中.

定义 3.1 设存在 $f_0 \in \mathbb{R}$, S 是可行集 Φ 的非空子集, 满足 $f(x) = f_0$, $\forall x \in S$, 令 γ 是正的常数. 称 γ 阶增长条件在 S 上成立, 若存在常数 $c > 0$, 和 S 的邻域 N, 满足对所有的 $x \in \Phi \cap N$, 下述不等式成立

$$f(x) \geqslant f_0 + c[\text{dist}(x, S)]^\gamma. \tag{3.4}$$

尤其, 若 $\gamma = 1$, 称一阶增长条件成立 (first order growth condition), 若 $\gamma = 2$, 称二阶增长条件成立.

很显然, 若存在 $\gamma > 0$, γ 增长条件于 S 上成立, 则 S 是 (P) 上的局部最优解集. 尤其, 若 $S = \{x_0\}$ 是单点集, 条件 (3.4) 具有下述形式

$$f(x) \geqslant f(x_0) + c\|x - x_0\|^\gamma, \quad \forall x \in \Phi \cap N, \tag{3.5}$$

其中 N 是 x_0 的邻域.

我们还考虑下述意义的“接近最优” (nearly optimal) 解.

定义 3.2 对 $\varepsilon \geqslant 0$, 称 $\bar{x}$ 是 (P) 的 ε 最优解, 若 $\bar{x} \in \Phi$ 且 $f(\bar{x}) \leqslant \inf_{x \in \Phi} f(x) + \varepsilon$.

3.1 一阶最优性条件

3.1.1 Lagrange 乘子

这一节讨论由 (3.1) 给出的问题 (P) 的 Lagrange 乘子形式的一阶必要性条件. 将 (P) 与其"标准"的参数化 (见 2.5.3 节) 相联系

$$(\mathrm{P}_y) \qquad \min_{x\in Q} f(x) \quad \text{s.t.} \quad G(x)+y\in K, \tag{3.6}$$

其相应的最优值函数是

$$v(y):=\inf\{f(x):\, x\in Q,\ G(x)+y\in K\}.$$

显然, $v(0)=\mathrm{val}(\mathrm{P})$. 问题 (P) 的 (共轭)对偶可以表示为下述形式 (见 (2.305))

$$(\mathrm{D}) \qquad \max_{\lambda\in Y^*}\left\{\inf_{x\in Q} L(x,\lambda)-\sigma(\lambda,K)\right\}. \tag{3.7}$$

通过研究问题 (P) 在可行点 x_0 处的如下形式的最优性条件开始讨论, 即存在 $\lambda\in Y^*$, 使得

$$x_0\in\arg\min_{x\in Q} L(x,\lambda), \quad \lambda\in N_K(G(x_0)). \tag{3.8}$$

注意到, 存在 λ 满足条件 $\lambda\in N_K(G(x_0))$ 可推出 $G(x_0)\in K$. 因此, 上述条件 (3.8) 可推出 x_0 是问题(P) 的可行点. 若集合 K 是凸锥, 则条件 $\lambda\in N_K(G(x_0))$ 等价于条件

$$G(x_0)\in K, \quad \lambda\in K^-, \quad \langle\lambda,G(x_0)\rangle=0. \tag{3.9}$$

给出注记, 除非问题 (P) 是凸的, 满足 (3.8) 的 λ 乘子集可能与将要定义的 Lagrange 乘子集是不同的.

由于集合 K 被假设为是凸的, 因此对应的指示函数$F(\cdot):=I_K(\cdot)$ 是凸的, 下述结果是定理 2.158 的一个简单的重新表述.

命题 3.3 *若 val(P)=val(D), $x_0\in X$, $\lambda\in Y^*$ 分别是 (P) 与 (D) 的最优解, 则条件 (3.8) 成立. 相反地, 若存在 x_0 与 λ, 条件 (3.8) 成立, 则 x_0 是 (P) 的最优解, λ 是 (D) 的最优解, 且 (P) 与 (D) 间没有对偶间隙.*

条件 val(P) =val(D) 可视为确保如果 x_0, λ 分别是原始与对偶问题的最优解, 则最优性条件 (3.8) 成立的约束规范. 然而这一条件是不容易验证的, 需要确保这样的乘子 λ 存在的更直接的约束规范. (P) 是平静的如果它的最优值 val(P) 是有限的且 $v(y)$ 在 $y=0$ 处是次可微的. 下述结果是 2.5.1 节给出的对偶理论的结论, 尤其可参看定理 2.142.

定理 3.4 (i) 如果 (P) 是平静的, 则问题 (P) 与 (D) 间不存在对偶间隙, 可行点 $x_0 \in \Phi$ 是问题 (P) 的最优解的充分必要条件是存在 $\lambda \in Y^*$, 满足条件 (3.8).

(ii) 相反的, 如果存在 x_0 与 λ 满足条件 (3.8), 则 (P) 是平静的.

(iii) 如果 (P) 是平静的, x_0 是 (P) 的最优解, 则满足最优性条件 (3.8) 的所有乘子 λ 的集合 Λ_0 是非空的凸的, 它与对偶问题(D) 的最优解集是重合的, 因此对 (P) 的任何最优解, 这一集合均是相同的.

我们现在考虑凸情况, 即 f 是凸函数, K 是闭凸集, G[①] 关于集合 $(-K)$ (见定义 2.163) 为凸的映射的情况, 则对任何 $\lambda \in N_K(y)$, $y \in K$, $L(\cdot,\lambda)$ 是 X 上的凸函数. 所以, 此种情况 $L(\cdot,\lambda) + I_Q(\cdot)$ 是凸函数, 条件 (3.8) 等价于

$$0 \in \partial_x(L(\cdot,\lambda) + I_Q(\cdot))(x_0), \quad \lambda \in N_K(G(x_0)). \tag{3.10}$$

若 f 与 G 还是连续的, 由 Moreau-Rockafellar 定理 (2.168 定理)(见注 2.169)

$$\partial_x(L(\cdot,\lambda) + I_Q(\cdot))(x_0) = \partial_x L(x_0,\lambda) + \partial_x I_Q(x_0),$$

进一步, $\partial_x I_Q(x_0) = N_Q(x_0)$. 结果, 系统 (3.10) 等价于

$$0 \in \partial_x L(x_0,\lambda) + N_Q(x_0), \quad \lambda \in N_K(G(x_0)). \tag{3.11}$$

定义 3.5 在问题 (P) 是凸的情况, 把满足条件 (3.11) 的向量 $\lambda \in Y^*$ 称为 Lagrange 乘子.

可见, 在凸的情况, Lagrange 乘子的集合与满足条件 (3.8) 的乘子 λ 的集合 Λ_0 是相同的.

若 (P) 是凸的, val(P) 有限, 或者空间 Y 是有限维的, 或者 K 具有非空内部, 则 (P) 是平静的充分必要条件是对所有的 $d \in Y$, $v'(0,d) > -\infty$(见命题 2.160).

由于这里设函数 $f(x)$ 与 $G(x)$ 是连续的, 集合 Q 是非空闭的, 函数 $f(x)+I_Q(x)$ 与 $I_K(G(x)+y)$ 是下半连续的正常的. 由定理 2.165 得, 若问题(P) 是凸的, val(P) 是有限的, 则正则性条件

$$0 \in \mathrm{int}\{G(Q) - K\} \tag{3.12}$$

(或等价地, (P_y) 对充分接近 $0 \in Y$ 的所有 y 的相容性) 等价于 $v(\cdot)$ 于 $0 \in Y$ 处是连续的. 后者可推出 $v(y)$ 在 $y = 0$ 处是次可微的, 因而问题 (P) 是平静的. 下述结果是定理 2.165 的结论.

定理 3.6 设问题 (P) 是凸的, x_0 是 (P) 的最优解, 正则性条件 (3.12) 成立, 则 Lagrange 乘子集合 Λ_0 是 Y^* 的非空的凸的有界的弱 * 紧致的子集, 它对 (P) 中的任何最优点均是相同的.

① 原著中为 $G(x)$.

若映射 G 是连续可微的, 相对于 $(-K)$ 是凸的, 则对任意 $x_0 \in \Phi$, 正则条件 (3.12) 等价于 Robinson 约束规范

$$0 \in \operatorname{int}\{G(x_0) + DG(x_0)(Q - x_0) - K\}. \tag{3.13}$$

现在考虑光滑的而不必是凸的情形. 即设函数 $f : X \to \mathbb{R}$ 与映射 $G : X \to X$ 是连续可微的. 导出一阶必要性最优条件的第一步是线性化 (P).

引理 3.7 *设 x_0 是 (P) 的局部最优解, 设 $f(\cdot)$ 与 $G(\cdot)$ 在 x_0 处是连续可微的. 则下述成立:*

(i) 点 $h = 0$ 是下述问题的最优解

$$\min_{h\in X} Df(x_0)h \quad \text{s.t.} \quad h \in T_\Phi(x_0). \tag{3.14}$$

(ii) 若 Robinson 约束规范(3.13) 成立, 则 $h = 0$ 是下述线性化问题的最优解

$$\min_{h\in X} Df(x_0)h \quad \text{s.t.} \quad h \in T_Q(x_0), \quad DG(x_0)h \in T_K(G(x_0)). \tag{3.15}$$

证明 (i) 令 $h \in T_\Phi(x_0)$. 由 (外) 切锥的定义, 存在序列 $t_n \downarrow 0$ 及 $x_n = x_0 + t_n h + o(t_n)$ 满足 $x_n \in \Phi$. 因为 x_0 是 (P) 的局部最优解, 有

$$0 \leqslant \lim_{n\to\infty} \frac{f(x_n) - f(x_0)}{t_n} = Df(x_0)h,$$

结果有 $h = 0$ 是问题 (3.14) 的最优解.

(ii) 由推论 2.91, 若 (3.13) 成立, 则 Φ 在 x_0 点处的内切集与外切集相等, 且等于问题 (3.15) 的可行集. 结合结论 (i), 证得该结论. □

以后提到问题 (P) 在 x_0 处的线性化问题时, 均指问题 (3.15). 注意到在结论 (ii) 的证明中, 只是为了证明 $T_\Phi(x_0)$ 与线性化问题 (3.15) 的可行集相等, 我们用到了 Robinson 约束规范. 在一些情况下, 可通过不同的方法来验证它, 比如, 对于线性约束问题, 上述结论仍然是成立的, 见 3.4.3 节.

在可微的情况, 一阶最优性条件可以表示为下述形式

$$-D_x L(x_0, \lambda) \in N_Q(x_0), \quad \lambda \in N_K(G(x_0)). \tag{3.16}$$

上述条件的第一个对应着 (在可微的凸情形是等价的)(3.11) 的第一个条件.

定义 3.8 *称 $\lambda \in Y^*$ 是问题 (P)(在点 x_0 处) 的 Lagrange 乘子, 若它满足条件 (3.16), 我们用 $\Lambda(x_0)$ 记 x_0 处所有 Lagrange 乘子的集合.*

注意到 (3.16) 中的第一个条件是 x_0 为 $L(\cdot, \lambda)$ 在 Q 上的一个极小点的必要条件. 所以有 $\Lambda_0 \subset \Lambda(x_0)$, 其中 Λ_0 是满足 (3.8) 的乘子集合. 对非凸问题这一包含关

系可能是严格的. 对于凸的可微的问题, $\Lambda(x_0)$ 与满足 (3.11) 的 λ 的集合是重合的, 因此, 上述定义与凸情形的 Lagrange 乘子的定义 3.5 没有矛盾.

注意, 法锥$N_Q(x_0)$ 与 $N_K(G(x_0))$ 非空当且仅当 $x_0 \in Q$, $G(x_0) \in K$. 因此, 由满足 (3.16) 的 Lagrange 乘子 λ 的存在性可推出点 x_0 的可行性. 还有, 若 $Q = X$, 则 $N_Q(x_0) = \{0\}$, 从而此时 (3.16) 中的第一个条件为 $D_xL(x_0,\lambda) = 0$. 若 K 是一凸锥, 则条件 (3.16) 可表示为

$$-D_xL(x_0,\lambda) \in N_Q(x_0), \quad G(x_0) \in K; \quad \lambda \in K^-, \quad \langle \lambda, G(x_0)\rangle = 0, \tag{3.17}$$

其中后三个条件等价于条件 $\lambda \in N_K(G(x_0))$.

由于 $[T_Q(x_0)]^- = N_Q(x_0)$, $[T_K(G(x_0))]^- = N_K(G(x_0))$, 线性化问题 (3.15) 的对偶可以表示为下述形式 (见 2.5.6 节)

$$\max_{\lambda} \ 0 \quad \text{s. t.} \quad -D_xL(x_0,\lambda) \in N_Q(x_0), \quad \lambda \in N_K(G(x_0)). \tag{3.18}$$

可见, 对偶问题 (3.18) 的可行集 (也是最优解集) 与 Lagrange 乘子集合 $\Lambda(x_0)$ 是相同的, 问题 (3.18) 有有限的最优值 (等于 0) 当且仅当集合 $\Lambda(x_0)$ 非空.

线性化问题 (3.15) 是凸的, 这一问题的约束规范(3.12) 为下述形式

$$0 \in \operatorname{int}\{DG(x_0)T_Q(x_0) - T_K(G(x_0))\}. \tag{3.19}$$

注意到, 由于 $(Q - x_0) \subset T_Q(x_0)$, $K - G(x_0) \subset T_K(G(x_0))$, 由原始问题的约束规范, 即 Robinson 约束规范 (3.13), 可推出 (3.19). 进一步, 若空间 X 与 Y 是有限维的, 则条件 (3.13) 与 (3.19) 是等价的. 这在命题 2.97 对 $Q = X$ 的情况已经给出证明, 推广到非空闭凸集 Q 的情况是直接的.

定理 3.9　*设 x_0 是问题 (P) 的局部最优解, f, G 是连续可微的, 且 Robinson 约束规范 (3.13) 成立. 则 Lagrange 乘子集 $\Lambda(x_0)$ 是 Y^* 中的非空的凸的有界的弱 * 紧致子集.*

证明　由引理 3.7, 在 Robinson 约束规范(3.13) 成立的前提下, $h = 0$ 是线性化问题 (3.15) 的最优解, 最优值是 0. 进一步, (3.13) 可推出 (3.19), 因而它是确保相应的最优值函数在 0 处是连续的线性化问题(3.15) 的约束规范. 由对偶理论 (见 2.5.3 节), 这可推出 (3.15) 与其对偶 (3.18) 间没有对偶间隙, 且 (3.18) 的最优解集是 Y^* 的非空的凸的有界的弱 * 紧致子集. 由于 (3.18) 的最优解集即 $\Lambda(x_0)$, 结论得证. □

集合

$$C(x_0) := \{h \in T_Q(x_0) : DG(x_0)h \in T_K(G(x_0)),\ Df(x_0)h \leqslant 0\} \tag{3.20}$$

被称为问题 (P) 在点 x_0 处的临界锥. 临界锥 $C(x_0)$ 表示的是线性化问题 (3.15) 在 x_0 处对最优性不能提供任何信息的那些方向, 在研究二阶最优性条件时尤其重要.

显然, $h=0$ 是线性化问题 (3.15) 的最优解的充分必要条件是对所有的 $h\in C(x_0)$ 有 $Df(x_0)h=0$. 此种情形, 临界锥与 (3.15) 的最优解集是重合的. 注意到, 即使 $h=0$ 不是线性化问题 (3.15) 的最优解, 上述临界锥 (3.20) 也是有意义的.

命题 3.10 令 $x_0\in\Phi$ 是 (P) 的一个可行点. 若 $\Lambda(x_0)$ 是非空的, 则 $h=0$ 是线性化问题 (3.15) 的最优解, 且

$$C(x_0)=\{h\in T_Q(x_0): DG(x_0)h\in T_K(G(x_0)),\ Df(x_0)h=0\}.$$

若有 $Q=X$, 则对每一 $\lambda\in\Lambda(x_0)$, 有

$$C(x_0)=\{h\in X: DG(x_0)h\in T_K(G(x_0)),\ \langle\lambda, DG(x_0)h\rangle=0\}.$$

证明 若 $\Lambda(x_0)$ 非空, 则对偶问题 (3.18) 有最优值 0, 因此原始问题 (3.15) 具有非负的最优值. 因为 $h=0$ 是可行的, 目标函数对应的值是 0, 这表明 $h=0$ 是 (3.15) 的最优解, 所以 $C(x_0)$ 的公式是成立的.

若还有 $Q=X$, 对每一 $\lambda\in\Lambda(x_0)$, $h\in X$ 满足 $DG(x_0)h\in T_K(G(x_0))$, 得

$$0=D_xL(x_0,\lambda)h=Df(x_0)h+\langle\lambda, DG(x_0)h\rangle,$$

因此, $Df(x_0)h=0$ 当且仅当 $\langle\lambda, DG(x_0)h\rangle=0$. 这完成了证明. □

我们已经看到, 有两个不同条件可使得 $h=0$ 是线性化问题 (3.15) 的最优解. 由引理 3.7, 一个条件是 $T_\Phi(x_0)$ 与线性化问题的可行集重合, 或由命题 3.10, 若 Lagrange 乘子集合是非空的. 这两个条件互不推出, 如下述例子所示.

例 3.11 考虑问题

$$\min_{x\in\mathbb{R}} x^2 \quad \text{s.t.} \quad x^2\leqslant 0,$$

此问题有唯一最优解 $x_0=0$. 我们有 $T_\Phi(x_0)$ 是 $\{0\}$, 它与线性化问题的可行集是不同的, 后者是 $\mathbb{R}$, 而 $\Lambda(x_0)=\mathbb{R}_+$ 是非空的.

例 3.12 考虑具有形式 (3.1) 的最优化问题, 其中 $X:=\mathbb{R}^2$, $Q:=X$, $Y:=\mathbb{R}^3, K:=\{(y_1,y_2,y_3)\in\mathbb{R}^3: y_3^2\geqslant y_1^2+y_2^2, y_3\geqslant 0\}, G(x_1,x_2):=(x_1,x_2,x_1), f(x_1,x_2):=x_2$, 该问题的可行集是 $\Phi=\{(x_1,x_2): x_2=0\}$. 有 $x_0:=(0,0)$ 是该问题的最优解, x_0 处线性化问题与原问题 (P) 是重合的. 另一方面, 若 $\lambda=(\lambda_1,\lambda_2,\lambda_3)$ 是问题的 Lagrange 乘子, 则它满足 $\lambda_1+\lambda_3=0, \lambda_2=-1$. 然而, 这样的 λ 不属于锥 K 的极锥. 实际上, 锥 K 的极锥是 $K^-=-K=\{(y_1,y_2,y_3)\in\mathbb{R}^3: y_3^2\geqslant y_1^2+y_2^2, y_3\leqslant 0\}$. 由于 $\lambda_1+\lambda_3=0, \lambda_1^2=\lambda_3^2, \lambda=(\lambda_1,\lambda_2,\lambda_3)$ 不会满足 $\lambda_3^2\geqslant\lambda_1^2+1$. 这表明此例子中的 Lagrange 乘子集合 $\Lambda(x_0)$ 是空集.

3.1.2 广义 Lagrange 乘子

这一节设 $f(x)$ 与 $G(x)$ 是连续可微函数, 函数

$$L^g(x,\alpha,\lambda) := \alpha f(x) + \langle \lambda, G(x) \rangle \tag{3.21}$$

称为问题 (P) 的广义 Lagrange 函数(generalized Lagrangian), 其中 $(x,\alpha,\lambda) \in X \times \mathbb{R} \times Y^*$. 显然, $\alpha = 1$, 广义 Lagrange 函数即 Lagrange 函数 $L(x,\lambda)$. 在可行点 x_0 处相应的广义一阶最优性条件可表示为下述形式

$$-D_x L^g(x_0,\alpha,\lambda) \in N_Q(x_0), \quad \lambda \in N_K(G(x_0)), \quad \alpha \geqslant 0, \quad (\alpha,\lambda) \neq (0,0). \tag{3.22}$$

定义 3.13 称 $(\alpha,\lambda) \in \mathbb{R} \times Y^*$ 是广义的Lagrange 乘子, 若它满足最优性条件 (3.22).

用 $\Lambda^g(x_0)$ 记满足最优性条件 (3.22) 的广义 Lagrange 乘子 $(\alpha,\lambda) \in \mathbb{R} \times Y^*$ 的集合. 注意条件 $(\alpha,\lambda) \neq (0,0)$ 是本质性的, 因为对于 $(\alpha,\lambda) = (0,0)$, (3.22) 中的其他条件是平凡成立的. 称广义Lagrange 乘子 $(\alpha,\lambda) \in \Lambda^g(x_0)$ 是奇异的(singular), 若 $\alpha = 0$. 若 $(0,\lambda)$ 是奇异的乘子, 也称 λ 为奇异的乘子. 奇异乘子集合记为

$$\Lambda^s(x_0) := \{\lambda \in Y^* : (0,\lambda) \in \Lambda^g(x_0)\}. \tag{3.23}$$

命题 3.14 设乘子集 $\Lambda(x_0)$ 是非空的. 则奇异 Lagrange 乘子集合 $\Lambda^s(x_0)$ 与 $\{0\}$ 构成 $\Lambda(x_0)$ 的回收锥.

证明 若 $(0,\bar{\lambda})$ 是奇异的 Lagrange 乘子, $\lambda \in \Lambda(x_0)$, 则 $\lambda + t\bar{\lambda} \in \Lambda(x_0), \forall t \geqslant 0$, 因此 $\bar{\lambda}$ 属于 $\Lambda(x_0)$ 的回收锥. 相反地, 设 μ 属于 $\Lambda(x_0)$ 的回收锥. 则对 $\lambda \in \Lambda(x_0)$ 对 $\forall t > 0, \lambda + t\mu \in \Lambda(x_0)$, 由于 $N_Q(x_0)$ 是锥, 有

$$[DG(x_0)]^*\mu + t^{-1}[DG(x_0)]^*\lambda + t^{-1}Df(x_0) \in -N_Q(x_0).$$

因为 $N_Q(x_0)$ 是闭集, 取 $t \to \infty$ 时上式的极限, 由上述包含关系得 $[DG(x_0)]^*\mu \in -N_Q(x_0)$. 类似地, 对 $\forall y \in T_K(G(x_0)), \langle \mu, y \rangle \leqslant 0$, 从而 $\mu \in N_K(G(x_0))$. 于是, 若还有 $\mu \neq 0$, 则 $(0,\mu)$ 是奇异的Lagrange 乘子. □

下述例子表明, 即使线性化问题 (3.15) 具有 $h = 0$ 为其最优解的情况, 也可能发生问题 (3.15) 与 (3.18) 具有对偶间隙, 且不存在广义 Lagrange 乘子满足 (3.22) 的情况.

例 3.15 设 $X = C[0,1], Y = L_2[0,1], Q = X, K = \{0\} \subset Y, G : X \to Y$ 是自然的嵌入映射, 即 G 认同 $x(\cdot) \in C[0,1]$ 为 $x(\cdot) \in L_2[0,1]$. 因空间 $L_2[0,1]$ 是 Hilbert 空间, 我们可认同 Y^* 为 Y. 有可行集 $\Phi = \{0\}$, 映射 G 是线性的, $G(X)$ 是 Y 中的稠密子集, $T_K(0) = \{0\}, N_K(0) = Y$. 考虑线性泛函 $f \in X^*$, 它定义为 $f(x) = x(a)$,

其中 a 是 $x(\cdot) \in C[0,1]$ 在区间 $[0,1]$ 中的不动点. 这一泛函表示质量 1 在点 a 处的原子度量(atomic measure of mass one at the point a). 这个例子中, 问题是线性的, 它与其线性化问题重合. 因为 $\Phi = \{0\}$, 显然 0 是这一问题的唯一最优解. 另一方面, 不难看到泛函 f 不属于伴随算子$G^* : Y^* \to X^*$ 的值域. 有相应的对偶问题具有空的可行集, 因此其最优值是 $-\infty$. 即原始与对偶问题间存在无穷的对偶间隙. 因为 $G(X)$ 在 Y 中是稠密的, $DG(x_0)^*\lambda = 0$ 可推出 $\lambda = 0$. 所以, 不存在奇异的 Lagrange 乘子, 因此, 此例中不存在广义 Lagrange 乘子.

现在考虑锥

$$\mathcal{Z}(x_0) := DG(x_0)[\mathcal{R}_Q(x_0)] - \mathcal{R}_K(G(x_0)). \tag{3.24}$$

由命题 2.95(可见 (2.194) 与 (2.195)), Robinson 约束规范(3.13) 等价于条件 $\mathcal{Z}(x_0) = Y$, 注意到一阶最优性条件 (3.16) 与 (3.22) 可视为线性化问题 (3.15) 的最优性条件. $\mathrm{ri}(S)$ 记为凸集合 S 的相对内部(见定义 2.16).

命题 3.16 *设 x_0 是 (P) 的局部最优解, 考虑 (3.24) 定义的锥 $\mathcal{Z}(x_0)$. 则下述结论成立:*

(i) *奇异的Lagrange 乘子存在当且仅当 $\mathrm{cl}[\mathcal{Z}(x_0)] \neq Y$.*

(ii) *再设锥 $\mathcal{Z}(x_0)$ 具有非空的相对内部. 则或者 $\mathcal{Z}(x_0) = Y$(即 Robinson 约束规范成立), 或存在奇异的 Lagrange 乘子.*

证明 设 $\mathrm{cl}[\mathcal{Z}(x_0)] \neq Y$, 即存在 $\bar{y} \notin \mathrm{cl}[\mathcal{Z}(x_0)]$. 则由于 $\mathcal{Z}(x_0)$ 是凸集, 由分离定理 2.14, $\mathcal{Z}(x_0)$ 与 $\bar{y}$ 可被非零的线性泛函 $\mu \in Y^*$ 分离, 即 $\langle \mu, y\rangle \leqslant \langle \mu, \bar{y}\rangle, \forall y \in \mathcal{Z}(x_0)$. 因为 $\mathcal{Z}(x_0)$ 是锥, 则 $\forall y \in \mathcal{Z}(x_0)$ 有 $\langle \mu, y\rangle \leqslant 0$, 从而 $\mu \in [\mathcal{Z}(x_0)]^-$. 进一步, 由 (2.31),

$$[\mathcal{Z}(x_0)]^- = [DG(x_0)(\mathcal{R}_Q(x_0))]^- \cap [-\mathcal{R}_K(G(x_0))]^-. \tag{3.25}$$

因此对 $\overline{\lambda} = -\mu$ 有 $\overline{\lambda} \in N_K(G(x_0)), DG(x_0)^*\overline{\lambda} \in -N_Q(x_0)$. 于是得 $(0, \overline{\lambda})$ 满足 (3.22), 它是奇异的广义 Lagrange 乘子. 相反地, 若 $(0, \overline{\lambda})$ 满足 (3.22), 则由 (3.25), $\overline{\lambda} \in [-\mathcal{Z}(x_0)]^-$. 因此若 $\mathrm{cl}[\mathcal{Z}(x_0)] = Y$, 有 $[-\mathcal{Z}(x_0)]^- = \{0\}$, 从而 $\overline{\lambda} = 0$. 这完成了 (i) 的证明.

性质 (ii) 可以类似证明. 即若 $\mathrm{cl}[\mathrm{Sp}\ [\mathcal{Z}(x_0)] \neq Y$, 则由 (i), 存在奇异的 Lagrange 乘子. 若 $\mathrm{cl}[\mathrm{Sp}\ [\mathcal{Z}(x_0)] = Y$, 则由非空相对内部的假设, $\mathcal{Z}(x_0)$ 在 Y 中有非空内部. 此种情况, 若存在 $\bar{y} \notin \mathcal{Z}(x_0)$, 则根据分离定理 2.13, $\mathcal{Z}(x_0)$ 与 $\bar{y}$ 可被非零的线性泛函 $\mu \in Y^*$ 分离. 可像 (i) 的证明那样继续下去. □

有如下的观察. 由命题 3.14 与 3.16(i) 得, 若 $\mathcal{Z}(x_0)$ 在 Y 中是稠密的, $\Lambda(x_0)$ 是非空的, 则 $\Lambda(x_0)$ 的回收锥是 $\{0\}$. 然而, 也可能发生 $\Lambda(x_0)$ 无界的情况 (见例 2.43 与例 3.21). 若存在奇异的 Lagrange 乘子, Lagrange 乘子集合 $\Lambda(x_0)$ 是非空的, 则 $\Lambda(x_0)$ 是无界的. 因为奇异的 Lagrange 乘子构成 $\Lambda(x_0)$ 的回收锥, 这一结论可立即

得到. 因此, 命题 3.16(ii) 给出条件, 在这些条件之下, $\Lambda(x_0)$ 的非空与有界性可推出 Robinson 约束规范(3.13).

命题 3.17 设 Lagrange 乘子集合 $\Lambda(x_0)$ 是非空有界的且锥 $\mathcal{Z}(x_0)$ 具有非空相对内部. 则 Robinson 约束规范 (3.13) 成立.

例 2.167 表明, 条件 $\mathrm{ri}(\mathcal{Z}(x_0)) \neq \varnothing$ 在上述命题中是本质性的. 即在无穷维空间的情形, Lagrange 乘子集非空且有界, 但 Robinson 约束规范可能不成立的情况是可能发生的.

命题 3.16(ii) 的另一个有意思的结论是关于广义 Lagrange 乘子存在性的下述结果.

命题 3.18 设 x_0 是 (P) 的局部最优解, 锥 $\mathcal{Z}(x_0)$ 具有非空的相对内部. 则存在广义的 Lagrange 乘子满足 (3.22).

现在给出 $\mathcal{Z}(x_0)$ 的相对内部非空的一些充分条件. Y 的线性子空间 Y_1 称为具有 (有限) 余维数 p, 若存在 p, 及线性独立的线性泛函 $a_1, \cdots, a_p \in Y^*$ 满足

$$Y_1 = \{y \in Y : \langle a_i, y\rangle = 0, i = 1, \cdots, p\}. \tag{3.26}$$

命题 3.19 设 $Q = X$, 在点 x_0 处的下述条件之一成立: (i) 集合 K 具有非空内部; (ii) $DG(x_0)$ 是映上的; (iii) Y 是有限维空间; (iv) 空间 Y 的子空间 $DG(x_0)X$ 具有有限的余维数. 则 $\mathcal{Z}(x_0)$ 的相对内部非空.

证明 给定 $y \in \mathrm{int}(K)$, 有 $d := y - G(x_0)$ 属于 $\mathcal{R}_K(G(x_0))$ 的内部, 因此有 $-d$ 属于 $\mathcal{Z}(x_0)$ 的内部. 这证得 (i). 因为假设 (ii) 与 (iii) 是 (iv) 的特殊情形, 只需证当 (iv) 成立, 结论是成立的. 置 $Y_1 := DG(x_0)X$, 令 $a_1, \cdots, a_p \in Y^*$ 满足 (3.26). 定义 $A \in \mathcal{L}(Y, \mathbb{R}^p)$ 满足 $Ay := (\langle a_1, y\rangle, \cdots, \langle a_p, y\rangle)$, 令 $y_0 := G(x_0), E := A(\mathcal{R}_K(y_0))$. 则 E 是 $\mathbb{R}^p$ 中的凸锥, 因而具有非空相对内部, 设 E 的维数是 q, $0 \leqslant q \leqslant p$. 在 $q < p$ 的情况, 可以选取 a_i 的 p 个线性组合, 如被记为 $b_1, \cdots, b_p$, 满足 b_i 是线性无关的, 且 $\langle b_i, k\rangle = 0, \forall k \in \mathcal{R}_K(y_0), i = q+1, \cdots, p$. 为记号的方便, 可设 $b_i = a_i, \forall i$. 令 $Y_2 := \{y \in Y : \langle a_i, y\rangle = 0, i = q+1, \cdots, p\}$, 用 A_q 记 A 的前 q 个分量. 显然, $\mathcal{Z}(x_0) \subset Y_2$. 设 $k_0 \in \mathcal{R}_K(y_0)$ 满足 Ak_0 属于 E 的相对内部. 我们论断 $-k_0$ 属于 $\mathcal{Z}(x_0)$ 相对于闭空间 Y_2 的内部. 事实上, 对于充分接近于 $-k_0$ 的 $y \in Y_2$, 则存在 $k' \in \mathcal{R}_K(y_0)$ 满足 $-A_q k' = A_q y$. 因此, $A(y + k') = 0$, 这等价于 $y + k' \in DG(x_0)X$. 这得到 $y = (y + k') - k'$ 属于 $\mathcal{Z}(x_0)$. □

下述例子说明, 在 K 具有空的内部的情形, 即使空间 X 是有限维的, 也可能发生广义 Lagrange 乘子不存在的情形.

例 3.20 令 $Q = X = \mathbb{R}, Y = L_2[0,1], K \subset L_2[0,1]$ 是几乎处处非正值函数构成的锥. 对 $x \in \mathbb{R}$, 定义 $f(x) := x$, $G(x)$ 是 $L_2[0,1]$ 中对应函数 $y(t) = t - x, t \in [0,1]$

的等价类. 显然, 这里的可行集是 $\Phi = \{x : x \geqslant 1\}$, $x_0 = 1$ 是最优解. 进一步, $T_K(G(x_0)) = Y$, 因而 $N_K(G(x_0)) = \{0\}$. 在此例中不存在广义 Lagrange 乘子.

下面的例子说明, 可能发生这样的情况, 即 Lagrange 乘子集无界, 不存在奇异的Lagrange 乘子.

例 3.21 设 $Q = X = \mathbb{R}, Y = l_2$. 考虑序列 $\bar{y} = (\bar{a_i}), \bar{a_i} := i^{-1}, i = 1, 2, \cdots$, 映射 $G : \mathbb{R} \to l_2$ 定义为 $G(x) := x\bar{y}$. 令 $K := \{(a_i) \in l_2 : a_i \geqslant 0, i = 1, 2, \cdots\}$, $f(x) = x$. 显然 $\Phi = \{x : x \geqslant 0\}$, $x_0 = 0$ 是对应问题 (P) 的最优解. 由于 $G(x_0) = 0$, 有 $\mathcal{R}_K(G(x_0)) = T_K(G(x_0)) = K$. 进一步, 在此例中不难验证 $DG(x_0)X - K$ 在 Y 中是稠密的. 由命题 3.16(i) 得, 此问题没有一奇异的 Lagrange 乘子 (这也可以直接验证). 另一方面, 考虑 $\lambda_k = (b_i^k)$, 其中若 $i \neq k$, 有 $b_i^k = 0$, 若 $i = k$, 则 $b_i^k := k$. 可直接验证, $\lambda_k, k = 1, 2, \cdots$ 是 Lagrange 乘子, $\|\lambda_k\| = k$. 所以在此例中 Lagrange 乘子集是无界的.

3.1.3 Ekeland 变分原理

即使空间 X 是有限维的, 由 (3.1) 定义的最优化问题 (P) 也可能没有最优解. 然而, 只要 (P) 的最优值是有限的, 对最优化过程而言, 对任意 $\varepsilon > 0$, 问题总有 ε 最优解 $\bar{x}$. 下述结果表明, 可构造另一接近 $\bar{x}$ 的 ε 最优解, 它是微小扰动的目标函数的极小点.

定理 3.22 (Ekeland 变分原理) *设 (E, ρ) 是完备的度量空间, $f : E \to \mathbb{R} \cup \{+\infty\}$ 是下半连续函数. 设 $\inf\limits_{e \in E} f(e)$ 是有限的, 对给定的 $\varepsilon > 0, \bar{e} \in E$ 是 f 的 ε 最优解, 即 $f(\bar{e}) \leqslant \inf\limits_{e \in E} f(e) + \varepsilon$. 则对任何 $k > 0$, 存在一点 $\hat{e} \in E$ 满足 $\rho(\bar{e}, \hat{e}) \leqslant k^{-1}$,*

$$f(\hat{e}) \leqslant f(\bar{e}) - \varepsilon k \rho(\bar{e}, \hat{e}), \tag{3.27}$$

$$f(\hat{e}) - \varepsilon k \rho(e, \hat{e}) < f(e), \quad \forall e \in E, e \neq \hat{e}. \tag{3.28}$$

证明 考虑如下定义的多值函数$M : E \to 2^E$,

$$M(e) := \{e' : f(e') + k\varepsilon\rho(e, e') \leqslant f(e)\}.$$

不难看到 $M(\cdot)$ 是自反的(reflexive), 即 $e \in M(e)$, 传递的(transitive), 即由 $e' \in M(e)$ 可推出 $M(e') \subset M(e)$. 考虑如下定义的函数 $v : \mathrm{dom} f \to \mathbb{R}$,

$$v(e) := \inf\{f(e') : e' \in M(e)\}.$$

有 $\inf\limits_E f \leqslant v(e) \leqslant f(e)$, 且

$$\varepsilon k \rho(e, e') \leqslant f(e) - v(e), \quad \forall e' \in M(e). \tag{3.29}$$

由于 $f(\bar{e}) - v(\bar{e}) \leqslant f(\bar{e}) - \inf\limits_{E} f \leqslant \varepsilon$, 得到

$$k\rho(\bar{e}, e) \leqslant 1, \quad \forall e \in M(\bar{e}). \tag{3.30}$$

结果, $M(\bar{e})$ 的直径(即两点距离的上确界) 小于或等于 $2k^{-1}$.

考虑满足下面关系的序列 $\{e_n\}$,

$$e_1 = \bar{e}, e_{n+1} \in M(e_n) \text{ 且 } f(e_{n+1}) \leqslant v(e_n) + \varepsilon 2^{-n}$$

(由 $v(\cdot)$ 的定义, 这样的序列是存在的). 因为 $M(e_{n+1}) \subset M(e_n)$(由 $M(\cdot)$ 的传递性), 得到 $v(e_n) \leqslant v(e_{n+1})$, 再由 $v(e) \leqslant f(e)$, 得到

$$v(e_{n+1}) \leqslant f(e_{n+1}) \leqslant v(e_n) + \varepsilon 2^{-n} \leqslant v(e_{n+1}) + \varepsilon 2^{-n},$$

因此得 $0 \leqslant f(e_{n+1}) - v(e_{n+1}) \leqslant \varepsilon 2^{-n}$. 结合 (3.29) 有, 对任何 $e \in M(e_{n+1})$, $k\rho(e_{n+1}, e) \leqslant 2^{-n}$, 从而当 $n \to \infty$ 时, $M(e_{n+1})$ 的直径收敛于零. 再由 $e_{n+1} \in M(e_n), M(e_{n+1}) \subset M(e_n)$, 得 $\{e_n\}$ 是 Cauchy 序列, 由空间 E 的完备性, $\{e_n\}$ 收敛到某一点 $\hat{e} \in \operatorname{dom}(f)$, 因 $M(e_n)$ 的直径收敛于 0, $\bigcap\limits_{n=1}^{\infty} M(e_n) = \{\hat{e}\}$. 因 $\{e_n\}$ 被包含在 $M(\bar{e})$ 中且 $M(\bar{e})$ 是闭的, 由 (3.30) 得到 $k\rho(\bar{e}, \hat{e}) \leqslant 1$, 根据 $M(\bar{e})$ 的定义, $f(\hat{e}) + \varepsilon k\rho(\bar{e}, \hat{e}) \leqslant f(\bar{e})$, 从而 (3.27) 成立. 再由 $M(\hat{e}) \subset M(e_n), \forall n, \operatorname{diam}(M(e_n)) \to 0$, 从而 $M(\hat{e}) = \{\hat{e}\}$, 可推出 (3.28). □

注意到, 上述定理中条件 (3.27) 表明 $\hat{e}$ 是 f 在 E 上的 ε 最优解, 条件 (3.28) 表明, $\hat{e}$ 是“扰动”函数 $f(\cdot) + \varepsilon k\rho(\cdot, \hat{e})$ 在 E 上的唯一的极小点. 尤其, 取 $k = \varepsilon^{-\frac{1}{2}}$, 对 f 的任何 ε 最优解 $\bar{e}$, 另一个 ε 最优解 $\hat{e}$ 存在, 满足 $\rho(\bar{e}, \hat{e}) \leqslant \varepsilon^{\frac{1}{2}}$ 且 $\hat{e}$ 是函数 $f(\cdot) + \varepsilon^{\frac{1}{2}}\rho(\cdot, \hat{e})$ 的极小点.

现在考虑由 (3.1) 给出的最优化问题 (P). 设 f 与 G 是连续可微的. 有可行集 Φ 是 Banach 空间 X 的闭子集, 因而可视为由 X 的范数引导的距离下的完备的度量空间. 结果, 若假设最优值 val(P) 是有限的, 则 Ekeland 变分原理可用于问题 (P).

设 $\bar{x}$ 是 (P) 的 ε 最优解, 其中 $\varepsilon > 0$. 则由 Ekeland 变分原理, 取 $k = \varepsilon^{-\frac{1}{2}}$, 存在 (P) 的 ε 最优解 $\hat{x}$ 满足 $\|\bar{x} - \hat{x}\| \leqslant \varepsilon^{\frac{1}{2}}$, 且 $\hat{x}$ 是下述函数

$$f_\varepsilon(x) := f(x) + \varepsilon^{\frac{1}{2}}\|x - \hat{x}\|$$

在集合 Φ 上的极小点. 设 Robinson 约束规范在点 $\hat{x}$ 处成立. 则类似于引理 3.7 的证明方式, 不难证明 $h = 0$ 是线性化问题

$$\min_{h \in X} Df(\hat{x})h + \varepsilon^{\frac{1}{2}}\|h\| + I_A(h) \quad \text{s. t.} \quad DG(\hat{x})h \in T_K(G(\hat{x})) \tag{3.31}$$

的最优解, 其中 $I_A(\cdot)$ 是集合 $A := T_Q(\hat{x})$ 的指示函数. (3.31) 中的项 $I_A(h)$ 对应着约束 $h \in T_Q(\hat{x})$.

上述优化问题是凸的, 因此定理 3.4 的最优性条件可用于 $h = 0$. 考虑问题 (3.31) 的 Lagrange 函数

$$\mathcal{L}_\varepsilon(h, \lambda) := Df(\hat{x})h + \varepsilon^{\frac{1}{2}}\|h\| + I_A(h) + \langle \lambda, DG(\hat{x})h \rangle.$$

注意到, 因为范数函数 $\psi(h) := \|h\|$ 是单位球 B_{X^*} 的支撑函数, 它在 $h = 0$ 处的次微分是 B_{X^*}, 指示函数 $I_A(h)$ 在 $h = 0$ 处的次微分是 $[T_Q(\hat{x})]^- = N_Q(\hat{x})$. 因此, 对固定的 $\lambda, \mathcal{L}_\varepsilon(\cdot, \lambda)$ 在 $h = 0$ 处的次微分可写为如下的形式

$$\partial \mathcal{L}_\varepsilon(0, \lambda) = Df(\hat{x}) + [DG(\hat{x})]^*\lambda + N_Q(\hat{x}) + \varepsilon^{\frac{1}{2}} B_{X^*}.$$

由 $h = 0$ 是 $\mathcal{L}_\varepsilon(\cdot, \lambda)$ 的极小点的充要条件是 $0 \in \partial \mathcal{L}_\varepsilon(0, \lambda)$, 得 $0 \in \underset{h \in X}{\operatorname{argmin}} \, \mathcal{L}_\varepsilon(h, \lambda)$ 当且仅当

$$\operatorname{dist}(-Df(\hat{x}) - [DG(\hat{x})]^*\lambda, N_Q(\hat{x})) \leqslant \varepsilon^{\frac{1}{2}}.$$

下面的结果是 Ekeland 变分原理的一个结论.

定理 3.23 *设 $\bar{x}$ 是 (P) 的 ε 最优解, 其中 $\varepsilon \geqslant 0, f$ 与 G 是连续可微的, Robinson 约束规范在所有的 $x \in \overline{B}(\bar{x}, \varepsilon^{\frac{1}{2}})$ 处均成立. 则存在 (P) 的另一 ε 最优解 $\hat{x}$ 及 $\lambda \in Y^*$ 满足 $\|\bar{x} - \hat{x}\| \leqslant \varepsilon^{\frac{1}{2}}$ 且*

$$\operatorname{dist}(-D_x L(\hat{x}, \lambda), N_Q(\hat{x})) \leqslant \varepsilon^{\frac{1}{2}}, \quad \lambda \in N_K(G(\hat{x})). \tag{3.32}$$

证明 若 $\varepsilon = 0$, 最优性条件 (3.32) 与定理 3.9 中相应的最优性条件重合. 若 $\varepsilon > 0$, 由上面的讨论, 结果可由 Ekeland 变分原理得到. □

定理 3.23 表明, 假设问题 (P) 的最优值是有限的, 约束规范成立, 最优化问题 (P) 总有 "接近最优" 的解, 这一解处一阶最优性条件 "几乎成立". 注意到, 若 $Q = X$, 则对任意的 $x \in X$, 有 $N_Q(x) = \{0\}$, 因此条件 (3.32) 具有下述形式

$$\|D_x L(\hat{x}, \lambda)\| \leqslant \varepsilon^{\frac{1}{2}}, \quad \lambda \in N_K(G(\hat{x})). \tag{3.33}$$

3.1.4 一阶充分条件

在某些相当特殊的情况, 可以给出可行点成为局部极小点的一阶充分条件. 这一节考虑由 (3.1) 给出的问题 (P), 设 $f(x)$ 与 $G(x)$ 是连续可微的. 注意比较下述的充分条件与引理 3.7 中的一阶必要条件. 对可行点 $x_0 \in \Phi$ 及 $\eta \geqslant 0$, 定义集合

$$\Upsilon_\eta(x_0) := \{h \in T_Q(x_0) : \operatorname{dist}(DG(x_0)h, T_K(G(x_0))) \leqslant \eta\|h\|\}. \tag{3.34}$$

注意到 $\Upsilon_\eta(x_0)$ 是一闭锥, 对 $\eta=0$, 它与线性化问题 (3.15) 的可行集是重合的, 即

$$\Upsilon_0(x_0)=\{h\in T_Q(x_0):DG(x_0)h\in T_K(G(x_0))\}. \tag{3.35}$$

引理 3.24　设 x_0 是问题 (P) 的可行点. 则

(i) 如果存在常数 $\alpha>0,\eta>0$ 满足

$$Df(x_0)h\geqslant\alpha\|h\|,\quad \forall h\in\Upsilon_\eta(x_0), \tag{3.36}$$

那么在 x_0 处有一阶增长条件成立.

(ii) 若正则性条件

$$DG(x_0)(T_Q(x_0))-T_K(G(x_0))=Y \tag{3.37}$$

成立, 对某一 $\alpha>0$ 与 $\eta=0$ 时的 (3.36) 成立, 则在 x_0 处一阶增长条件成立.

(iii) 设 Robinson 约束规范在 x_0 处成立. 置 $\eta=0$. 则 $\alpha>0$ 存在使得 (3.36) 成立的充分必要条件是 x_0 处一阶增长条件成立.

证明　(i) 考虑可行点 $x_0+h\in\Phi$. 因为 $Q\subset x_0+T_Q(x_0)$, 有 $h\in T_Q(x_0)$. 进一步, 由于 $G(x_0+h)\in K\subset G(x_0)+T_K(G(x_0))$, 由 $h\mapsto G(x_0+h)$ 的一阶展式得

$$\operatorname{dist}(DG(x_0)h,T_K(G(x_0)))=o(\|h\|). \tag{3.38}$$

因此, 对于充分接近于零的 h, 有 $h\in\Upsilon_\eta(x_0)$, 由此由 (3.36) 得

$$f(x_0+h)=f(x_0)+Df(x_0)+o(\|h\|)\geqslant f(x_0)+\alpha\|h\|+o(\|h\|),$$

这就证得 (i).

(ii) 由 (3.37), 根据稳定性定理 (定理 2.87) 得, 若 $h\in T_Q(x_0)$ 满足 (3.38), 则存在 $h'\in\Upsilon_0(x_0)$ 满足 $\|h-h'\|=o(\|h\|)$. 由 f 的一阶 Taylor 展式, 如上可完成证明.

(iii) 因为 $\mathcal{R}_K(G(x_0))\subset T_K(G(x_0))$, 由 Robinson 约束规范 (3.13) 可得正则条件 (3.37)(见命题 2.95). 因此由 (ii) 可得必要性条件 (即 x_0 处的一阶增长条件成立). 只需证明若 x_0 满足 Robinson 约束规范且一阶增长条件成立, 则 $\eta=0$ 时 (3.36) 成立. 令 $h\in T_Q(x_0)$ 满足 $DG(x_0)h\in T_K(G(x_0))$. 则

$$\operatorname{dist}(x_0+h,\,Q)+\operatorname{dist}(G(x_0+h),K)=o(\|h\|).$$

由命题 2.89, Φ 是度量正则的, 因此存在 $x(h)\in\Phi$ 满足 $x(h)=x_0+h+o(\|h\|)$. 由常数 $c>0$ 的一阶增长条件得

$$Df(x_0)h=Df(x_0)(x(h)-x_0)+o(\|h\|)\geqslant c\|x(h)-x_0\|+o(\|h\|)\geqslant c\|h\|+o(\|h\|),$$

则得到 (3.36). □

注意, 对 $\eta=0$, 条件 (3.36) 可解释为线性化问题 (3.15) 的一阶增长条件. 注意到, 对合适选取的 x_0 的邻域 N, 一阶增长的常数可选取小于 (3.36) 中的常数 α 的任何正数.

若空间 X 是有限维的, 由紧致性不难证明, 一阶增长条件的充分条件是

$$Df(x_0)h>0, \forall h\in T_Q(x_0)\setminus\{0\} \text{ 满足 } DG(x_0)h\in T_K(G(x_0)). \tag{3.39}$$

进一步, 若 X 是有限维的, Robinson 约束规范 (3.13) 成立, 则上述条件是一阶增长条件的充分必要条件. 注意, 条件 (3.39) 意味着 $h=0$ 是线性化问题 (3.15) 的唯一的最优解. 在无穷维空间的情形, 条件 (3.39) 不能保证 x_0 的局部最优性, 如例 3.25 所示.

例 3.25 令 $X=Y=l_2, Q=X, K:=\{(x_i):x_i\geqslant 0, \forall i=1,2,\cdots\}, G(x)=x, f(x)=\langle\alpha,x\rangle-\langle x,x\rangle$, 其中 $\alpha=(\alpha_i)\in l_2, \forall i=1,2,\cdots,\alpha_i>0$. 显然 $Df(0)=\alpha, Df(0)h>0$ 对任何非零的 $h\in K=T_K(0)$ 成立. 因为 $DG(x)$ 是映上的, 约束规范的最强形式成立. 然而, $x_0=0$ 不是局部最优的. 事实上, 考虑点列 $x^k=(x_i^k)\in l_2, k=1,2,\cdots$, 满足若 $k=i$ 则 $x_i^k=2\alpha_k$, 且若 $k\neq i$ 则 $x_i^k=0$. 显然, $x^k\in K, \|x^k\|=2\alpha_k$, 从而 $x^k\to 0$. 进一步, 对任何 k, $f(x^k)=2(\alpha_k)^2-4(\alpha_k)^2<0$, 因此 $x_0=0$ 不是局部最优的.

可以给出与充分条件 (3.36) 等价的各种不同的条件. 锥 $\Upsilon_0(x_0)$ 表示的是线性化问题的可行集.

命题 3.26 对 $\eta=0$, 条件 (3.36) 等价于

$$-Df(x_0)\in \operatorname{int}([\Upsilon_0(x_0)]^-). \tag{3.40}$$

证明 条件 (3.40) 意味着, 存在 $r>0$,

$$-Df(x_0)+rB_{X^*}\subset [\Upsilon_0(x_0)]^-.$$

等价地, 对任何 $h\in\Upsilon_0(x_0)$ 及所有的 $x^*\in B_{X^*}$,

$$\langle -Df(x_0)+rx^*,h\rangle\leqslant 0.$$

这等价于

$$Df(x_0)h\geqslant r\sup_{x^*\in B_{X^*}}\langle x^*,h\rangle=r\|h\|,$$

即等价于 $\eta=0$ 时的 (3.36). □

计算锥 $\Upsilon_0(x_0)$ 的极锥需要下述结果.

引理 3.27 设 X 与 Y 是 Banach 空间, $A : X \to Y$ 是连续的线性算子, $C \subset X$ 与 $K \subset Y$ 是闭凸锥. 则

$$A^*(K^-) + C^- \subset [A^-(K) \cap C]^-, \tag{3.41}$$

且若 $A(C) - K = Y$, 则

$$A^*(K^-) + C^- = [A^-(K) \cap C]^-. \tag{3.42}$$

证明 因为 A 是线性的且连续的, $M := A^-(K)$ 是 X 中闭凸锥. 令 $y^* \in K^-$. 则对任何 $x \in M$ 有 $Ax \in K$ 且

$$\langle A^*y^*, x\rangle = \langle y^*, Ax\rangle \leqslant 0.$$

这表明 $A^*y^* \in M^-$, 因而有 $A^*(K^-) \subset M^-$. 由 (2.32) 得

$$M^- + C^- \subset (M \cap C)^-,$$

因此证得 (3.41).

现在设 $A(C) - K = Y$. 对 $x^* \in X^*$, 考虑问题

$$\min_{h\in C} \langle x^*, h\rangle \quad \text{s. t.} \quad Ah \in K. \tag{3.43}$$

它的对偶 (见 (2.347)) 是

$$\max_{y^* \in K^-} 0 \quad \text{s. t.} \quad x^* + A^*y^* \in -C^-. \tag{3.44}$$

在正则条件 $A(C) - K = Y$ 之下, (3.43) 与 (3.44) 问题间没有对偶间隙 (见 2.5.6 节). 这两个问题的公共最优值或是 0 或 $-\infty$. 问题 (3.43) 的最优值是 0 当且仅当 $-x^* \in (M \cap C)^-$, 问题 (3.44) 的最优值是 0 当且仅当存在 $y^* \in K^-$ 满足 $x^* + A^*y^* \in -C^-$. 这意味着, 若 $-x^* \in (M \cap C)^-$, 则存在 $y^* \in K^-, c \in C^-$ 满足 $-x^* = A^*y^* + c$. 即 $(M \cap C)^- \subset A^*(K^-) + C^-$, 因而 (3.42) 成立. □

现在设正则条件 (3.37) 成立. 则由引理 3.27 得

$$[\Upsilon_0(x_0)]^- = [DG(x_0)]^* N_K(G(x_0)) + N_Q(x_0).$$

因此, 包含关系 (3.40) 与 $\eta = 0$ 时的 (3.36) 等价于

$$-Df(x_0) \in \operatorname{int}\{[DG(x_0)]^* N_K(G(x_0)) + N_Q(x_0)\}. \tag{3.45}$$

尤其, 若 $Q = X$, 则 (3.45) 具有下述形式

$$-Df(x_0) \in \operatorname{int}\{[DG(x_0)]^* N_K(G(x_0))\}. \tag{3.46}$$

显然, (3.46) 右端的锥包含在空间 $[DG(x_0)]^*Y^*$ 中. 所以, 若空间 Y 是有限维的, 因而 Y^* 是有限维的, 如 $\dim Y = m$, 则 $\dim[DG(x_0)]^*Y^* \leqslant m$. 于是, 若 $\dim X > \dim Y$(尤其, 若 X 是无穷维空间, Y 是有限维空间), 则条件 (3.46) 不会成立. 这表明, 像 (3.36) 或 (3.46) 这样的一阶充分条件只能在某些特殊情况下成立. 我们将会看到 (见 3.4.3 节), 对最优解集 S 非空的线性规划问题, 一阶增长条件在 S 上总是成立的. 这一结果与上述讨论不是矛盾的, 因为若 $\dim X > \dim Y$, 对应的线性规划问题的最优解集 S 非空, 则 S 是凸的且无界的, 因此不能包含孤立点.

3.2 二阶必要性条件

3.2.1 二阶切集

继续二阶分析需要一个描述集合的可能的曲率的工具. 这一节通过研究二阶切集来处理这一问题, 通过类似于 (一阶) 切锥的方式定义二阶切集. 设 X 与 Y 是 Banach 空间, S 是 X 的 (可能非凸的) 子集. 回顾 Painlevé-Kuratowski 意义下的上、下集合极限的定义 2.5.2.

定义 3.28 *集合极限*

$$T_S^{i,2}(x,h) := \liminf_{t\downarrow 0} \frac{S - x - th}{\frac{1}{2}t^2}, \tag{3.47}$$

$$T_S^2(x,h) := \limsup_{t\downarrow 0} \frac{S - x - th}{\frac{1}{2}t^2} \tag{3.48}$$

分别称为 S 在点 x 沿方向 h 的内二阶切集与外二阶切集.

同样, 这两个切集可表示为如下形式

$$T_S^{i,2}(x,h) = \left\{w \in X : \operatorname{dist}\left(x + th + \frac{1}{2}t^2 w,\ S\right) = o(t^2), t \geqslant 0\right\}, \tag{3.49}$$

$$T_S^2(x,h) = \left\{w \in X : \exists t_n \downarrow 0, \operatorname{dist}\left(x + t_n h + \frac{1}{2}t_n^2 w,\ S\right) = o(t_n^2)\right\}. \tag{3.50}$$

由上述定义, 显然有 $T_S^{i,2}(x,h) \subset T_S^2(x,h)$. 若 $T_S^{i,2}(x,h)$ 是非空的, 则 $\operatorname{dist}(x+th,\ S) = O(t^2), t \geqslant 0$. 所以, 只有 $h \in T_S^i(x)$, 内二阶切集$T_S^{i,2}(x,h)$ 才可能非空. 类似地, 若 $T_S^2(x,h)$ 是非空的, 则存在序列 $t_n \downarrow 0$ 满足 $\operatorname{dist}(x + t_n h,\ S) = O(t_n^2)$. 因此, 只有 $h \in T_S(x)$, 外二阶切集 $T_S^2(x,h)$ 才可能是非空的. 因为下、上集合极限是闭的, 则两个集合 $T_S^{i,2}(x,h)$ 与 $T_S^2(x,h)$ 是闭的.

若 S 是凸的, 距离函数 $\operatorname{dist}(\cdot, S)$ 是凸的. 由 (3.49) 得内二阶切集$T_S^{i,2}(x,h)$ 是凸的. 另一方面, 即使 S 是凸集, 外二阶切集$T_S^2(x,h)$ 也可能是非凸的 (见例 3.35).

距离函数 $\operatorname{dist}(\cdot, S)$ 是 Lipschitz 连续的 (模 1). 因此, 若当 $t \downarrow 0$, $w(t) \to w$ 且 $\operatorname{dist}\left(x + th + \frac{1}{2}t^2 w(t),\, S\right) = o(t^2)$, 则 $w \in T_S^{i,2}(x,h)$, 对外二阶切集$T_S^2(x,h)$ 也有类似的结果. 即 $T_S^{i,2}(x,h)$ 与 $T_S^2(x,h)$ 是闭的.

例 3.29 考虑闭凸集

$$S := \{(x_1, x_2) \in \mathbb{R}^2 : x_2 \geqslant |x_1|^{\frac{3}{2}}\},$$

点 $x_0 := (0,0)$ 与方向 $h := (1,0)$. 有 $T_S(x_0) = \{(x_1, x_2) : x_2 \geqslant 0\}, h \in T_S(x_0)$. 然而

$$\lim_{t \downarrow 0} \frac{\operatorname{dist}(x_0 + th,\, S)}{t^2} = \infty,$$

因此集合 $T_S^{i,2}(x_0,h)$ 与 $T_S^2(x_0,h)$ 均是空集.

我们也可从下述的观点来看二阶切集. 用 Σ 记收敛到 0 的正数序列 $\{t_n\}$ 的集合. 对序列 $\sigma := \{t_n\} \in \Sigma$, 定义下述 (内) 序列二阶切集(sequential second order tangent set)

$$T_S^{i,2,\sigma}(x,h) := \left\{w : \operatorname{dist}\left(x + t_n h + \frac{1}{2}t_n^2 w,\, S\right) = o(t_n^2)\right\}, \tag{3.51}$$

或等价地,

$$T_S^{i,2,\sigma}(x,h) = \liminf_{n\to\infty} \frac{S - x - t_n h}{\frac{1}{2}t_n^2}. \tag{3.52}$$

对任何 $\sigma \in \Sigma$, 集合 $T_S^{i,2,\sigma}(x,h)$ 是闭的, 且它在 S 是凸的情况是凸集. 显然, 取 $\sigma \in \Sigma$ 的所有的 $T_S^{i,2,\sigma}(x,h)$ 的交是集合 $T_S^{i,2}(x,h)$, 所有的 $T_S^{i,2,\sigma}(x,h)$ 的并是 $T_S^2(x,h)$.

设集合 S 由一个约束

$$S := \{x \in X : g(x) \leqslant 0\}$$

定义, 其中 $g(\cdot)$ 是凸的下半连续函数. 下述命题中, 将 S 的二阶切集与 $g(\cdot)$ 的二阶方向导数相联系. 因为 $g(x)$ 是凸函数, 其方向上图导数$g^{\downarrow}(x_0,\cdot)$ 存在且是凸的下半连续函数.

命题 3.30 设集合 S 定义为形式 $S := \{x \in X : g(x) \leqslant 0\}$, 其中 $g(\cdot)$ 是正常的凸的下半连续函数. 令 $g(x_0) = 0$ 且 $g^{\downarrow}(x_0,h) = 0$, 设存在 $\bar{x}$ 满足 $g(\bar{x}) < 0$(Slater 条件). 则

$$T_S^2(x_0,h) = \{w \in X : g_-^{\downarrow\downarrow}(x_0;h,w) \leqslant 0\}, \tag{3.53}$$

$$T_S^{i,2}(x_0,h) = \{w \in X : g_+^{\downarrow\downarrow}(x_0;h,w) \leqslant 0\}. \tag{3.54}$$

证明 先证明 (3.53) 成立. 考虑 $w \in T_S^2(x_0, h)$, 选取序列 $t_n \downarrow 0, w_n \to w$ 满足 $x_0 + t_n h + \frac{1}{2}t_n^2 w_n \in S$, 因而有 $g\left(x_0 + t_n h + \frac{1}{2}t_n^2 w_n\right) \leqslant 0$. 则

$$g_-^{\downarrow\downarrow}(x_0; h, w) \leqslant \frac{g\left(x_0 + t_n h + \frac{1}{2}t_n^2 w_n\right)}{\frac{1}{2}t_n^2} + o(1) \leqslant o(1).$$

从而有 $g_-^{\downarrow\downarrow}(x_0; h, w) \leqslant 0$.

相反地, 设 $g_-^{\downarrow\downarrow}(x_0; h, w) < 0$. 则存在 $t_n \downarrow 0, w_n \to w$, 有

$$g\left(x_0 + t_n h + \frac{1}{2}t_n^2 w_n\right) = \frac{1}{2}t_n^2 g_-^{\downarrow\downarrow}(x_0; h, w) + o(t_n^2),$$

对充分大的 $n, g\left(x_0 + t_n h + \frac{1}{2}t_n^2 w_n\right) < 0$. 结果

$$x_0 + t_n h + \frac{1}{2}t_n^2 w_n \in S,$$

即有 $w \in T_S^2(x_0, h)$.

现在设 $g_-^{\downarrow\downarrow}(x_0; h, w) = 0$, 则存在 $t_n \downarrow 0, w_n \to w$ 满足 $g\left(x_0 + t_n h + \frac{1}{2}t_n^2 w_n\right) = o(t_n^2)$. 给定 $\alpha > 0$, 置 $w_\alpha := w + \alpha(\bar{x} - x_0)$. 由 g 的凸性可得, 对充分小的 $t \geqslant 0$, 满足 $1 - \frac{1}{2}\alpha t^2 > 0$,

$$g\left(x_0 + th + \frac{1}{2}t^2 w_\alpha\right) \leqslant \left(1 - \frac{1}{2}\alpha t^2\right)\gamma(t, w) + \frac{1}{2}\alpha t^2 g(\bar{x}), \tag{3.55}$$

其中

$$\gamma(t, w) := g\left(x_0 + t\left(1 - \frac{1}{2}\alpha t^2\right)^{-1} h + \frac{1}{2}t^2\left(1 - \frac{1}{2}\alpha t^2\right)^{-1} w\right).$$

定义 t'_n 与 w'_n 满足关系: $t'_n\left(1 - \frac{1}{2}\alpha t'_n\right)^{-1} = t_n$, 即 $t'_n = t_n\left(1 + \frac{1}{2}\alpha t_n\right)^{-1}$ 与 $\left(1 - \frac{1}{2}\alpha t'_n\right)w'_n = w_n$. 则

$$\gamma(t'_n, w'_n) = g\left(x_0 + t_n h + \frac{1}{2}t_n^2 w_n\right) = o(t_n^2).$$

由于 $t'_n \downarrow 0, w'_n + \alpha(\bar{x} - x_0) \to w_\alpha$ 及 $g(\bar{x}) < 0$, 由 (3.55) 及对任意 $\alpha > 0$, 有

$$g_-^{\downarrow\downarrow}(x_0; h, w_\alpha) \leqslant \alpha g(\bar{x}) < 0,$$

因此有 $w_\alpha \in T_S^2(x_0, h)$. 因为 $T_S^2(x_0, h)$ 是闭的, 令 $\alpha \downarrow 0$, 得到 $w \in T_S^2(x_0, h)$. 这就完成了 (3.53) 的证明. □

由上述结果, 对任何序列 $\sigma = \{t_n\} \in \Sigma$, 集合

$$\limsup_{n\to\infty} \frac{S - x_0 - t_n h}{\frac{1}{2}t_n^2} \tag{3.56}$$

与集合 $\{w : \psi^\sigma(w) \leqslant 0\}$ 重合, 其中

$$\psi^\sigma(\cdot) := \mathrm{e} - \liminf_{n\to+\infty} \frac{g\Big(x_0 + t_n h + \frac{1}{2}t_n^2 \cdot\Big)}{\frac{1}{2}t_n^2}.$$

通过对 $\sigma \in \Sigma$ 取 $\psi^\sigma(\cdot)$ 上确界, 得到 $g_+^{\downarrow\downarrow}(x_0; h, \cdot)$, 通过取形式 (3.56) 的所有集合的交, 得 $T_S^{i,2}(x_0, h)$. 这样就得到 (3.54). □

若 $g(\cdot)$ 是凸的且在 x_0 处是连续的, 则 $g^\downarrow(x_0, \cdot) \equiv g'(x_0, \cdot)$, $g_-^{\downarrow\downarrow}(x_0; h, \cdot) \equiv g''_-(x_0; h, \cdot)$, $g_+^{\downarrow\downarrow}(x_0; h, \cdot) \equiv g''_+(x_0; h, \cdot)$. 进一步, 若 $g(x_0) < 0$, 即 x_0 是 S 的内部点, 因此 $T_S^{i,2}(x_0, h) = T_S^2(x_0, h) = X$. 若 $g(x_0) = 0$ 且 $g'(x_0, h) < 0$, 则对 $t > 0$ 充分小, $x_0 + th$ 是 S 的内部点, 有 $T_S^{i,2}(x_0, h) = T_S^2(x_0, h) = X$.

下述例子说明, 即使是对凸集, 也不像一阶切锥, 二阶内切集与二阶外切集可能是不同的.

例 3.31 考虑凸的分片线性函数 $y = \eta(x)$, 它在两个抛物线 $y = x^2$ 与 $y = 2x^2$ 间震荡. 如例 2.69 那样构造. 令 $S := \{(x_1, x_2) \in \mathbb{R}^2 : \eta(x_1) \leqslant x_2\}$, 即 S 是 η 的上图. 如例 2.69 所证明的那样, 所联系的函数 $g(x_1, x_2) := \eta(x_1) - x_2$ 在 $(0,0)$ 处沿方向 $h := (1, 0)$ 不是二阶方向可微的. 结果, 此例中相应的内二阶切集与外二阶切集是不同的. 事实上, 不难验证 (用公式 (3.53) 与 (3.54)) $T_S^{i,2}(0, h) = \{(x_1, x_2) : x_2 \geqslant 4\}$, 而 $T_S^2(0, h) = \{(x_1, x_2) : x_2 \geqslant 2\}$. 进一步, 对任何序列 $\sigma \in \Sigma$, 存在 $a \in [2, 4]$, 相反地, 对任何 $a \in [2, 4]$, 可以构造序列 $\sigma \in \Sigma$, 满足 $T_S^{i,2,\sigma}(0, h) = \{(x_1, x_2) : x_2 \geqslant a\}$.

定义 3.32 *称集合 S 在 $x \in S$ 处沿方向 $h \in T_S(x)$ 是二阶方向可微的, 若 $T_S^i(x) = T_S(x)$ 且 $T_S^{i,2}(x, h) = T_S^2(x, h)$.*

若 S 在 $x \in S$ 处沿方向 $h \in T_S(x)$ 是二阶方向可微的, 则对任何的 $\sigma \in \Sigma$, 有 $T_S^{i,2,\sigma}(x, h) = T_S^{i,2}(x, h)$. 由命题 3.30 得, 若 S 由凸约束定义, Slater 条件成立, 则 S 在 $x \in S$ 沿满足 $g^\downarrow(x, h) = 0$ 的方向 h 是二阶方向可微的充分必要条件是水平集$\{w : g_-^{\downarrow\downarrow}(x; h, w) \leqslant 0\}$ 与 $\{w : g_+^{\downarrow\downarrow}(x; h, w) \leqslant 0\}$ 重合. 尤其, 若函数 $g(\cdot)$ 是凸连续的且在 x 处沿方向 h 是二阶上图可微的, 则 S 是二阶方向可微的.

现在设 $S := G^{-1}(K) = \{x \in X : G(x) \in K\}$, 其中 K 是 Banach 空间 Y 的闭凸子集, $G : X \to Y$ 是二阶连续可微的映射. 下述的公式给出用 K 的二阶切集表示的计算 S 的二阶切集的法则. Σ 是收敛到 0 的正数序列的全体.

命题 3.33 设 $K \subset Y$ 是闭凸集, $G: X \to Y$ 是二阶连续可微的映射, $x_0 \in S := G^{-1}(K)$. 设 Robinson 约束规范 (2.178) 成立. 则对所有的 $h \in X$ 与任意的序列 $\sigma = \{t_n\} \in \Sigma$,

$$T_S^{i,2,\sigma}(x_0, h) = DG(x_0)^{-1}[T_K^{i,2,\sigma}(G(x_0), DG(x_0)h) - D^2G(x_0)(h,h)]. \tag{3.57}$$

证明 考虑点 $w \in T_S^{i,2,\sigma}(x_0, h)$, 令 $x_n := x_0 + t_n h + \frac{1}{2}t_n^2 w$ 是相应的抛物序列. 由 G 的二阶 Taylor 展开, 有

$$G(x_n) = G(x_0) + t_n DG(x_0)h + \frac{1}{2}t_n^2[DG(x_0)w + D^2G(x_0)(h,h)] + o(t_n^2). \tag{3.58}$$

因为 G 是连续的可微的, 因此是局部 Lipschitz 连续的, 及 $\operatorname{dist}(x_n, S) = o(t_n^2)$, 得到 $\operatorname{dist}(G(x_n), K) = o(t_n^2)$. 结合 (3.58) 得到

$$DG(x_0)w + D^2G(x_0)(h,h) \in T_K^{i,2,\sigma}(G(x_0), DG(x_0)h),$$

因此, (3.57) 的左端包含在 (3.57) 的右端. 相反的包含关系可通过与上述的相反的推证及应用稳定性定理 (定理 2.87) 得到. □

由命题 3.33 的假设及 (3.57) 得到

$$T_S^{i,2}(x_0, h) = DG(x_0)^{-1}[T_K^{i,2}(G(x_0), DG(x_0)h) - D^2G(x_0)(h,h)], \tag{3.59}$$

$$T_S^2(x_0, h) = DG(x_0)^{-1}[T_K^2(G(x_0), DG(x_0)h) - D^2G(x_0)(h,h)]. \tag{3.60}$$

注意到, 尽管集合 K 设为凸集, 集合 $S := G^{-1}(K)$ 不一定是凸集. 若映射 G 是 $(-K)$ 凸的, 则 S 是凸的.

命题 3.34 设 S 是凸集. 则对任意的 $x \in S, h \in T_S(x), \sigma = \{t_n\} \in \Sigma$. 下述包含关系成立:

$$T_S^{i,2,\sigma}(x, h) + T_{T_S(x)}(h) \subset T_S^{i,2,\sigma}(x, h) \subset T_{T_S(x)}(h). \tag{3.61}$$

证明 令 $w \in T_S^{i,2,\sigma}(x, h)$, 即 $x + t_n h + \frac{1}{2}t_n^2 w + o(t_n^2) \in S$. 考虑一点 $z \in S$. 因为 S 是凸的, $x \in S$, 则对任何 $\alpha > 0$ 及满足 $\alpha t_n \leqslant 1$ 的 t_n, 有 $x + \alpha t_n(z - x) \in S$. 将上述两个包含关系取权数为 $\left(1 - \frac{1}{2}\beta t_n\right)$ 与 $\frac{1}{2}\beta t_n$ 的凸组合, 其中 $\beta > 0$ $\Big($设 t_n 充分小使 $\frac{1}{2}\beta t_n < 1\Big)$, 得到

$$x + t_n h + \frac{1}{2}t_n^2\{w + \beta[\alpha(z - x) - h]\} + o(t_n^2) \in S,$$

因此有

$$w + \beta[\alpha(z - x) - h] \in T_S^{i,2,\sigma}(x, h).$$

由于 $z-x$ 是 $\mathcal{R}_S(x)$ 的任意元素, $T_S^{i,2,\sigma}(x,h)$ 是闭的, 得

$$w+\beta[T_S(x)-h]\subset T_S^{i,2,\sigma}(x,h),$$

因为 $\beta>0$ 是任意的, $T_S(x)$ 是凸的, 同样可得

$$w+T_{T_S(x)}(h)\subset T_S^{i,2,\sigma}(x,h),$$

即第一包含关系成立.

对凸集 $S, x\in S$, 包含关系 $S\subset x+T_S(x)$ 总是成立的. 得到

$$\operatorname{dist}\left(x+t_nh+\frac{1}{2}t_n^2w,\ S\right)\geqslant \operatorname{dist}\left(t_nh+\frac{1}{2}t_n^2w, T_S(x)\right),$$

因上述不等式的左端是 $o(t_n^2)$ 阶的, 得到 $\operatorname{dist}\left(h+\frac{1}{2}t_nw, T_S(x)\right)=o(t_n)$. 这表明 $\frac{1}{2}w\in T_{T_S(x)}(h)$, 因而由 $T_{T_S(x)}(h)$ 是锥, (3.61) 的第二包含关系成立. □

对凸集 S, 由 (3.61) 可得下述包含关系

$$T_S^{i,2}(x,h)+T_{T_S(x)}(h)\subset T_S^{i,2}(x,h)\subset T_{T_S(x)}(h), \tag{3.62}$$

$$T_S^2(x,h)+T_{T_S(x)}(h)\subset T_S^2(x,h)\subset T_{T_S(x)}(h). \tag{3.63}$$

则若 $0\in T_S^2(x,h)$, 有 $T_S^2(x,h)=T_{T_S(x)}(h)$. 进一步, 若 $0\in T_S^{i,2}(x,h)$, 即有 $\operatorname{dist}(x+th,\ S)=o(t^2)$, 这三个集合重合, 即

$$T_S^{i,2}(x,h)=T_S^2(x,h)=T_{T_S(x)}(h).$$

尤其, 若 $\mathcal{R}_S(x)=T_S(x)$, 如集合 S 是多面集, 则 $0\in T_S^{i,2}(x,h)$, 因而上述三个集合相同. 注意到, 若 $h\in T_S(x)$, 有 (见例 2.62)

$$T_{T_S(x)}(h)=\operatorname{cl}\{T_S(x)+[\![h]\!]\}. \tag{3.64}$$

注意, 由包含关系 (3.61) 可推出, 若 $T_S^{i,2,\sigma}(x,h)$ 是非空的, 则 $[T_S^{i,2,\sigma}(x,h)]^\infty$ 与锥 $T_{T_S(x)}(h)$ 相等, 同样地, 对 $T_S^{i,2}(x,h)$ 与 $T_S^2(x,h)$ 也有类似的结论.

设 $X_1,\cdots,X_n$ 是 Banach 空间, $S_i\subset X_i, i=1,\cdots,n$ 是闭集. 考虑空间 $X:=X_1\times\cdots\times X_n$, 它赋予范数 $\|x\|:=\sum_{i=1}^n\|x_i\|$, 其中 $x=(x_1,\cdots,x_n)\in X$, 集合 $S:=S_1\times\cdots\times S_n\subset X$. 则 $\operatorname{dist}(x,\ S)=\sum_{i=1}^n\operatorname{dist}(x_i,\ S_i)$, 因此

$$T_S^i(x)=T_{S_1}^i(x_1)\times\cdots\times T_{S_n}^i(x_n), \tag{3.65}$$

对 $h := (h_1, \cdots, h_n) \in X, \sigma \in \Sigma$,

$$T_S^{i,2,\sigma}(x,h) = T_{S_1}^{i,2,\sigma}(x_1,h_1) \times \cdots \times T_{S_n}^{i,2,\sigma}(x_n,h_n). \tag{3.66}$$

因此得到

$$T_S^{i,2}(x,h) = T_{S_1}^{i,2}(x_1,h_1) \times \cdots \times T_{S_n}^{i,2}(x_n,h_n) \tag{3.67}$$

及

$$T_S^2(x,h) \subset T_{S_2}^2(x_1,h_1) \times \cdots \times T_{S_n}^2(x_n,h_n). \tag{3.68}$$

进一步, 若集合 S_2 在 x_2 处沿方向 h_2 是二阶方向可微的, 即对任何序列 $\sigma \in \Sigma$, $T_{S_2}^{i,2}(x_2,h_2) = T_{S_2}^{i,2,\sigma}(x_2,h_2) = T_{S_2}^2(x_2,h_2)$, 则 (对 $n=2$)

$$T_S^2(x,h) = T_{S_1}^2(x_1,h_1) \times T_{S_2}^2(x_2,h_2). \tag{3.69}$$

如下面例子所示, (3.68) 的包含关系可能是严格的(即在 (3.68) 中等式不成立), 即使在 S_i 是凸集的情况, 外二阶切集可能是非凸的. 即类似于 (3.67) 的公式对外二阶切集是不成立的.

例 3.35 我们如下构造函数 $y = \alpha(x), x \in \mathbb{R}$, 它在两个抛物线 $y = x^2$ 与 $y = 2x^2$ 间振荡. 从点 $x_1 > 0$ 出发, 画抛物线 $y = 2x^2$ 的切线, 切点为 $(x_1, 2x_1^2)$. 此直线与抛物线 $y = x^2$ 在点 (a_1, a_1^2) 相交, $a_1 < x_1$. 画一条经过 (a_1, a_1^2) 的与 $y = 2x^2$ 相切的直线. 令 $(b_1, 2b_1^2)$ 是相应的切点, $0 < b_1 < a_1$, 令 $\alpha(\cdot)$ 为定义在 $[b_1, x_1]$ 上的相应的分片线性函数. 置 $x_2 := b_1 - \ell_1$, 其中 $\ell_1 > 0$ 满足 $x_2 > 0$, 对所有的 $x \in [x_2, b_1]$ 定义 $\alpha(x) := 2x^2$, 继续下去. 置 $\alpha(0) := 0$, 对负的 x 置 $\alpha(-x) := \alpha(x)$. 这样构造的函数 $\alpha(x)$ 是凸的, 它在区间 $[b_k, x_k]$ 上是分片线性的, 在区间 $[x_{k+1}, b_k]$ 上是二次的. 注意区间 $[x_{k+1}, b_k]$ 的长度 l_k 可以是任意的, 只要满足 $x_{k+1} > 0$ 即可.

令 $y = \beta(x)$ 是两个抛物线 $y = x^2$ 与 $y = 2x^2$ 间振荡的函数, 其构造是类似的, 相应的序列是 $\{x'_k\}$ 与 $\{b'_k\}$ 满足 $x'_k = b_k$, 即 $\beta(\cdot)$ 在 $\alpha(\cdot)$ 为分片线性的区间上是二次函数, $\beta(\cdot)$ 在 $\alpha(\cdot)$ 为二次函数的区间上是分片线性的. 用相同的方式, 满足这一性质的两个函数 $\beta(\cdot)$ 与 $\alpha(\cdot)$ 可用上述迭代过程同时构造.

定义集合 $S_1 := \operatorname{epi}\alpha$, $S_2 := \operatorname{epi}\beta$, $S := S_1 \times S_2$. 因为 α 与 β 是凸的, 这些集合是凸的. 对 $h_1 = h_2 := (1,0)$, 得到 (与例 3.31 比较)

$$T_{S_1}^{i,2}(0,h_1) = T_{S_2}^{i,2}(0,h_2) = \{(x_1,x_2) : x_2 \geqslant 4\},$$

$$T_{S_1}^2(0,h_1) = T_{S_2}^2(0,h_2) = \{(x_1,x_2) : x_2 \geqslant 2\}.$$

然而满足 $T_{S_j}^{2,\sigma_j}(0,h_j) = T_{S_j}^2(0,h_j), j = 1,2$ 的序列 $\sigma_1, \sigma_2 \in \Sigma$ 是不同的. 进一步, 由上述构造, 对 $h := (h_1, h_2)$ 得

$$T_S^2(0,h) = (T_{S_1}^{i,2}(0,h_1) \times T_{S_2}^2(0,h_2)) \cup (T_{S_1}^2(0,h_1) \times T_{S_2}^{i,2}(0,h_2)).$$

我们得到, 集合 $T_S^2(0,h)$ 是非空的且相应的包含关系 (3.68) 是严格的.

命题 3.36 设 $S_1,\cdots,S_n$ 是 Banach 空间 X 的闭凸集, $S:=S_1\cap\cdots\cap S_n$, 设存在点 $\bar{x}\in S_n$ 满足 $\bar{x}\in \mathrm{int}(S_i), i=1,\cdots,n-1$. 则

(i) 对 $x\in S$,

$$T_S(x)=T_{S_1}(x)\cap\cdots\cap T_{S_n}(x). \tag{3.70}$$

(ii) 对 $x\in S, h\in T_{S_1}(x)\cap\cdots\cap T_{S_n}(x)$ 及 $\sigma\in\Sigma$,

$$T_S^{i,2,\sigma}(x,h)=T_{S_1}^{i,2,\sigma}(x,h)\cap\cdots\cap T_{S_n}^{i,2,\sigma}(x,h). \tag{3.71}$$

证明 证 (ii), (i) 的证明类似. 考虑 X 到 $X^n:=X\times\cdots\times X$ 的映射 $G(x):=(x,\cdots,x)$, 令 $K:=S_1\times\cdots\times S_n\subset X^n$. 显然 $G^{-1}(K)=S$. 下证在上述假设之下, 对映射 G 的 Robinson 约束规范成立. 由命题的假设, 存在 $\varepsilon>0$ 满足 $\bar{x}+2\varepsilon B_X\subset S_1\cap\cdots\cap S_{n-1}$. 若 $u_1,\cdots,u_n\in\varepsilon B_X$, 令 $x=\bar{x}+u_n$, 则 $s_i:=x-u_i\in\bar{x}+2\varepsilon B_X\subset S_i, i=1,\cdots,n-1$. 因此, 若置 $s_n:=\bar{x}\in S_n$, 有 $u_i=x-s_i\in x-S_i, i=1,\cdots,n$, 因而 $[\varepsilon B_X]^n\subset G(X)-K$, 这证得 Robinson 约束规范成立.

由 (3.66) 得

$$T_K^{i,2,\sigma}(G(x),G(h))=T_{S_1}^{i,2,\sigma}(x,h)\times\cdots\times T_{S_n}^{i,2,\sigma}(x,h).$$

因为 G 是线性的, 由 (3.57) 得 (3.71). □

在上述命题的假设下, 有

$$T_S^{i,2}(x,h)=T_{S_1}^{i,2}(x,h)\cap\cdots\cap T_{S_n}^{i,2}(x,h), \tag{3.72}$$

$$T_S^2(x,h)\subset T_{S_1}^2(x,h)\cap\cdots\cap T_{S_n}^2(x,h). \tag{3.73}$$

注意到, (3.73) 中的包含关系可以是严格的 (如例 3.35).

命题 3.37 设 $S_1,\cdots,S_n$ 是 Banach 空间 X 中的闭 (但不一定是凸的) 子集, 令 $S:=\bigcup\limits_{k=1}^n S_k$. 则

$$T_S(x)=\bigcup_{k=1}^n T_{S_k}(x), \tag{3.74}$$

$$\bigcup_{k=1}^n T_{S_k}^i(x)\subset T_S^i(x), \tag{3.75}$$

$$T_S^2(x,h)=\bigcup_{k=1}^n T_{S_k}^2(x,h), \tag{3.76}$$

$$\bigcup_{k=1}^n T_{S_k}^{i,2}(x,h)\subset T_S^{i,2}(x,h). \tag{3.77}$$

证明 观察到

$$\mathrm{dist}(\cdot,\, S) = \min\{\mathrm{dist}(\cdot,\, S_k) : k = 1, \cdots, n\}. \tag{3.78}$$

即使某一 S_i 是空集, 此时相应的距离函数是 $+\infty$, 上面的公式也是成立的. 设对某一 $k \in \{1, \cdots, n\}, h \in T^i_{S_k}(x)$. 则由内切集的定义有 $\mathrm{dist}(x + th,\, S_k) = o(t), t \geqslant 0$. 结合 (3.78), 这表明 $\mathrm{dist}(x + th,\, S) = o(t), t \geqslant 0$, 从而 $h \in T^i_S(x)$. 这证得 (3.75). (3.77) 的证明是类似的.

现在设 $h \in T_{S_k}(x)$ 对某一 $k \in \{1, \cdots, n\}$ 成立. 这意味着存在序列 $t_n \downarrow 0$ 满足 $\mathrm{dist}(x + t_n h,\, S_k) = o(t_n)$. 由 (3.78) 有 $\mathrm{dist}(x + t_n h,\, S) = o(t_n)$, 从而 $h \in T_S(x)$. 相反地, 设 $h \in T_S(x)$. 则存在序列 $t_n \downarrow 0$ 满足 $\mathrm{dist}(x + t_n h,\, S) = o(t_n)$. 根据 (3.78), 存在 $k \in \{1, \cdots, n\}$ 及 $\{t_n\}$ 的子序列 $\{t_{n(m)}\}$ 满足 $\mathrm{dist}(x + t_{n(m)}h,\, S_k) = o(t_{n(m)})$①, 因此有 $h \in T_{S_k}(x)$. 这证得 (3.74). 公式 (3.76) 的证明类似. □

下述例子表明, (3.75) 与 (3.77) 中包含关系可能是严格的.

例 3.38 考虑 $\mathbb{R}$ 中的集合 $S_1 := \bigcup\limits_{k=0}^{\infty} [2^{-(2k+1)}, 2^{-2k}] \cup \{0\}$ 与 $S_2 := \bigcup\limits_{k=1}^{\infty} [2^{-2k}, 2^{-(2k-1)}] \cup \{0\}$. 有 $S := S_1 \cup S_2 = [0, 1]$ 且 $T^i_S(0) = T_S(0) = \mathbb{R}_+$. 另一方面, 不难验证 $T^i_{S_1}(0) = T^i_{S_2}(0) = \{0\}$, 因此相应的包含关系 (3.75) 是严格的. 进一步, $T^{i,2}_S(0, 1) = T^2_S(0, 1) = \mathbb{R}$, 集合 $T^{i,2}_{S_1}(0, 1)$ 与 $T^{i,2}_{S_2}(0, 1)$ 是空集, 因此包含关系 (3.77) 是严格的.

注意到, $T_{S_1}(0) = T_{S_2}(0) = \mathbb{R}_+, T^2_{S_1}(0, 1) = T^2_{S_2}(0, 1) = \mathbb{R}$, 因此等式 (3.74) 与 (3.76) 成立.

例 3.39 设集合 S 定义为下述形式

$$S := \{x \in X : g_i(x) = 0, i = 1, \cdots, q; g_i(x) \leqslant 0, i = q + 1, \cdots, p\}, \tag{3.79}$$

其中约束函数 $g_i : X \mapsto \mathbb{R}$ 是二阶连续可微函数. 注意到, 上述集合 S 可以表示为形式 $S = G^{-1}(K)$, 其中 $G(x) := (g_1(x), \cdots, g_p(x)) : X \to \mathbb{R}^p, K := \{0\} \times \mathbb{R}^{p-q}_-$. 令 $x_0 \in S$, 设 Mangasarian-Fromovitz 约束规范(2.191) 在 x_0 处成立. 由推论 2.91 得

$$T_S(x_0) = \{h \in X : Dg_i(x_0)h = 0, i = 1, \cdots, q, Dg_i(x_0)h \leqslant 0, i \in I(x_0)\}, \tag{3.80}$$

其中

$$I(x_0) := \{i : g_i(x_0) = 0, i = q + 1, \cdots, p\}$$

记为在 x_0 点起作用的不等式约束的集合.

集合 K 是多面集, 因而 K 的内外二阶切集是重合的, 对 $y \in K$, 有

$$T_K(y) = \{d \in \mathbb{R}^p : d_i = 0, i = 1, \cdots, q; d_i \leqslant 0, i \in \mathcal{I}(y)\},$$

①原著中为 $\mathrm{dist}(x + t_{n(m)}h,\, S) = o(t_{n(m)})$.

且对 $d \in T_K(y)$,

$$T_K^2(y,d) = \{w \in \mathbb{R}^p : w_i = 0, i = 1, \cdots, q; w_i \leqslant 0, i \in \mathcal{I}_1(y,d)\},$$

其中

$$\mathcal{I}(y) := \{i : y_i = 0, i = q+1, \cdots, p\}, \quad \mathcal{I}_1(y,d) := \{i \in \mathcal{I}(y) : d_i = 0\}.$$

用链式法则 (3.59) 与 (3.60), 得 $T_S^{i,2}(x_0,h) = T_S^2(x_0,h)$, 且对 $h \in T_S(x_0)$,

$$T_S^2(x_0,h) = \left\{ \begin{array}{r} w \in X : Dg_i(x_0)w + D^2 g_i(x_0)(h,h) = 0, \ i = 1, \cdots, q, \\ Dg_i(x_0)w + D^2 g_i(x_0)(h,h) \leqslant 0, \ i \in I_1(x_0,h) \end{array} \right\}, \tag{3.81}$$

其中

$$I_1(x_0,h) := \{i \in I(x_0) : Dg_i(x_0)h = 0\}.$$

例 3.40　计算 $p \times p$ 阶负半定矩阵构成的锥 $K \subset \mathcal{S}^p$ 的二阶切集可以用公式 (3.53). K 可以由一个凸约束定义为 $K = \{X \in \mathcal{S}^p : \lambda_{\max}(X) \leqslant 0\}$. 可以证明 (见例 4.145) 最大特征值函数 $\lambda_{\max}(\cdot) : \mathcal{S}^p \to \mathbb{R}$ 具有二阶方向导数, 可表示为

$$\lambda''_{\max}(A; H, W) = \lambda_{\max}(F^{\mathrm{T}} E^{\mathrm{T}} (W - 2H(A - \lambda_{\max}(A) I_p)^{\dagger} H) E F), \tag{3.82}$$

其中 $A, H, W \in \mathcal{S}^p$ 是对称矩阵, $A^{\dagger}$ 记 A 的 Moore-Penrose 伪逆, $E = [e_1, \cdots, e_s]$ 是 $p \times s$ 矩阵, 它的列 $e_1, \cdots, e_s$ 构成对应于最大特征值的 A 的特征向量空间的一组正交基, $F = [f_1, \cdots, f_r]$ 的列 $f_1, \cdots, f_r$ 构成对应 $s \times s$ 矩阵 $E^{\mathrm{T}} H E$ 最大特征值的特征向量空间的一组直交基.

现在设 $\lambda_{\max}(A) = 0, \lambda'_{\max}(A, H) = 0$. 回顾 $\lambda'_{\max}(A, H) = \lambda_{\max}(E^{\mathrm{T}} H E)$. Slater 条件显然成立. 因此, 有 $T_K^{i,2}(A,H) = T_K^2(A,H)$ 且

$$T_K^2(A,H) = \{W \in \mathcal{S}^p : F^{\mathrm{T}} E^{\mathrm{T}} W E F \preceq 2 F^{\mathrm{T}} E^{\mathrm{T}} H A^{\dagger} H E F\}. \tag{3.83}$$

注意到由于 $\lambda_{\max}(A) = 0$, 矩阵 A 是负半定的, $AE = 0, E^{\mathrm{T}} E = I_s, \mathrm{rank} A = p - s$,

$$A^{\dagger} = \sum_{i=1}^{p-s} \lambda_i(A)^{-1} a_i a_i^{\mathrm{T}}, \tag{3.84}$$

其中 $\lambda_i(A), i = 1, \cdots, p - s$ 是矩阵 A 的非零特征值, 且 a_i 是对应的直交的特征向量. 类似地, 由于 $\lambda_{\max}(E^{\mathrm{T}} H E) = 0$, 有 $E^{\mathrm{T}} H E \preceq 0, (E^{\mathrm{T}} H E) F = 0, F^{\mathrm{T}} F = I_r, \mathrm{rank}(E^{\mathrm{T}} H E) = s - r$.

设 $f(x)$ 是增广的实值 (不必是凸的) 函数. 则类似于命题 2.58 中的关于切锥的公式 (2.100) 与 (2.101), 有下述的 f 的上图的二阶切集的公式.

命题 3.41 设 $f(x): X \to \overline{\mathbb{R}}$ 是增广的实值函数, 在点 $x \in X$ 处取有限值. 则

$$T^{i,2}_{\text{epi}f}[(x, f(x)), (h, f^{\downarrow}_{+}(x,h))] = \text{epi} f^{\downarrow\downarrow}_{+}(x; h, \cdot), \tag{3.85}$$

$$T^{2}_{\text{epi}f}[(x, f(x)), (h, f^{\downarrow}_{-}(x,h))] = \text{epi} f^{\downarrow\downarrow}_{-}(x; h, \cdot) \tag{3.86}$$

在 $f^{\downarrow}_{-}(x,h)$ 与 $f^{\downarrow}_{+}(x,h)$ 有限的前提下是成立的.

关于集合的链式法则 (3.59) 与 (3.60) 可翻译为复合函数的相应的链式法则. 下述二阶链式法则可类似于命题 2.136 的证明加以证明. 若 f 是凸函数, 在 x_0 处取有限值, 则 $f^{\downarrow}(x_0, \cdot)$ 存在 (见 2.2.3 节).

命题 3.42 设 $G: X \to Y$ 是二阶连续可微映射, $f: Y \to \overline{\mathbb{R}}$ 是下半连续的凸函数, 在点 $y_0 := G(x_0)$ 处取有限值. 设 Robinson 约束规范成立

$$0 \in \text{int}\{G(x_0) + DG(x_0)X - \text{dom} f\}. \tag{3.87}$$

则假设 $f^{\downarrow}(y_0; DG(x_0)h)$ 是有限的, 就得到

$$(f \circ G)^{\downarrow\downarrow}_{-}(x_0; h, w) = f^{\downarrow\downarrow}_{-}(y_0; DG(x_0)h, DG(x_0)w + D^2G(x_0)(h,h)), \tag{3.88}$$

$$(f \circ G)^{\downarrow\downarrow}_{+}(x_0; h, w) = f^{\downarrow\downarrow}_{+}(y_0; DG(x_0)h, DG(x_0)w + D^2G(x_0)(h,h)). \tag{3.89}$$

例 3.43 设 A 是 X 的非空闭凸子集, $f(\cdot) := I_A(\cdot), x \in A$, 考虑集合 $K := \text{epi}\, I_A = A \times \mathbb{R}_+$. 则 $f^{\downarrow}(x, \cdot) := I_{T_A(x)}(\cdot)$(见例 2.67). 考虑向量 $h \in T_A(x)$. 不难看到

$$f^{\downarrow\downarrow}_{-}(x; h, w) = \begin{cases} 0, & \text{若 } w \in T^2_A(x,h), \\ +\infty, & \text{否则}. \end{cases} \tag{3.90}$$

进一步, 由 (3.67) 得

$$T^{2,i}_K((x,0),(h,\gamma)) = \begin{cases} T^{2,i}_A(x,h) \times \mathbb{R}, & \text{若 } \gamma > 0, \\ T^{2,i}_A(x,h) \times \mathbb{R}_+, & \text{若 } \gamma = 0, \\ \varnothing, & \text{若 } \gamma < 0. \end{cases} \tag{3.91}$$

由 (3.69) 得

$$T^{2}_K((x,0),(h,\gamma)) = \begin{cases} T^{2}_A(x,h) \times \mathbb{R}, & \text{若 } \gamma > 0, \\ T^{2}_A(x,h) \times \mathbb{R}_+, & \text{若 } \gamma = 0, \\ \varnothing, & \text{若 } \gamma < 0. \end{cases} \tag{3.92}$$

因此, 下述条件是等价的: (i) 集合 $K := \text{epi}\, I_A$ 在 $(x, 0)$ 处是二阶方向可微的; (ii) 集合 A 在 x 处是二阶方向可微的; (iii) 函数 I_A 在点 x 处是二阶上图可微的 ①.

① 原著中是 x_0.

3.2.2　二阶必要条件的一般形式

这一节讨论二阶必要条件. 为简化表述, 讨论下述形式问题 (P) 的二阶最优性条件(先是必要性的, 后面是充分性的)

$$\text{(P)} \qquad \min_{x\in X} f(x) \quad \text{s. t.} \quad G(x)\in K, \tag{3.93}$$

即设 $Q=X$ 的情况. 我们还给出如何把这样的二阶条件推广到 Q 是 X 的 (凸的闭的) 子集的方法. 设函数 f 与 G 是二次连续可微的, K 是 Y 的闭凸子集.

设 x_0 是 (P) 的局部 (极小点) 最优解. 考虑线性化问题

$$\min_{h\in X} Df(x_0)h \quad \text{s. t.} \quad DG(x_0)h\in T_K(G(x_0)), \tag{3.94}$$

临界锥是

$$C(x_0):=\{h\in X: DG(x_0)h\in T_K(G(x_0)), Df(x_0)h\leqslant 0\}. \tag{3.95}$$

若集合 (Lagrange 乘子集合)$\Lambda(x_0)$ 是非空的, 则临界锥的上述定义中, 条件 $Df(x_0)h\leqslant 0$ 可替换为 $Df(x_0)h=0$.

对 $h,w\in X$, 考虑路径$x(\cdot):\mathbb{R}_+\to X$, 它具有下述形式

$$x(t)=x_0+th+\frac{1}{2}t^2w+r(t), \tag{3.96}$$

满足 $r(t)=o(t^2)$. 则由 G 在 x_0 处的二阶 Taylor 展式得

$$G(x(t))=G(x_0)+tDG(x_0)h+\frac{1}{2}t^2[DG(x_0)w+D^2G(x_0)(h,h)]+o(t^2). \tag{3.97}$$

由外二阶切集(见 (3.50)) 定义, 存在 $t_n\downarrow 0$ 满足 $\operatorname{dist}(G(x(t_n)),K)=o(t_n^2)$ 当且仅当

$$DG(x_0)w+D^2G(x_0)(h,h)\in T_K^2(G(x_0),DG(x_0)h), \tag{3.98}$$

其中 $T_K^2(G(x_0),DG(x_0)h)$ 是 K 在 $G(x_0)$ 沿方向 $DG(x_0)h$ 的外二阶切集. 于是导致下述二阶必要条件.

引理 3.44　*设 x_0 是问题 (P) 的局部解且 Robinson 约束规范在 x_0 处是成立的. 则对每一 $h\in C(x_0)$ 及满足 (3.98) 的所有 $w\in X$, 成立*

$$Df(x_0)w+D^2f(x_0)(h,h)\geqslant 0. \tag{3.99}$$

证明　考虑 $h\in C(x_0)$ 及满足 (3.98) 的 w. 则存在 $t_n\downarrow 0$ 满足 $\operatorname{dist}(G(x(t_n)),K)=o(t_n^2)$. 因此, 由稳定性定理 (定理 2.87) 得, 可以取 (3.96) 中的项 $r(t_n)$ 使 $x(t_n)\in\Phi$, 即 $x(t_n)$ 是可行的, 且 $r(t_n)=o(t_n^2)$. 进一步, 由 f 在 x_0 处的二阶 Taylor 展式有

$$f(x(t_n))=f(x_0)+t_nDf(x_0)h+\frac{1}{2}t_n^2[Df(x_0)w+D^2f(x_0)(h,h)]+o(t_n^2), \tag{3.100}$$

由 $h \in C(x_0)$ 有 $Df(x_0)h = 0$. 因为 $x(t_n)$ 是可行的, 所以对 n 充分大有 $f(x(t_n)) \geqslant f(x_0)$, 因而得到 (3.99) 式. □

注意到在临界方向 h 的定义中, 条件 $DG(x_0)h \in T_K(G(x_0))$ 在必要条件 (3.99) 中被隐含地用到, 因为, 否则外二阶切集 $T_K^2(G(x_0), DG(x_0)h)$ 是空集. 也注意到, 这些必要条件可以用一最优化问题来描述. 即对任何 $h \in C(x_0)$, 下述问题

$$\begin{aligned} \min_{w \in X} \quad & Df(x_0)w + D^2 f(x_0)(h,h), \\ \text{s. t.} \quad & DG(x_0)w + D^2 G(x_0)(h,h) \in T_K^2(G(x_0), DG(x_0)h) \end{aligned} \tag{3.101}$$

的最优值是非负的.

因为在一般情况下, (凸集的) 二阶切集不是凸集, 上述问题一般来说不是凸问题. 然而, 通过考虑可行集的凸的子集, 我们可以把二阶必要条件 (3.101) 表示为下述的对偶形式.

定理 3.45 (二阶必要条件) *设 x_0 是 (P) 的局部最优解, Robinson 约束规范在 x_0 处成立. 则对每一 $h \in C(x_0)$, 对任意凸集合 $\mathcal{T}(h) \subset T_K^2(G(x_0), DG(x_0)h)$, 下述不等式成立*

$$\sup_{\lambda \in \Lambda(x_0)} \{D_{xx}^2 L(x_0, \lambda)(h,h) - \sigma(\lambda, \mathcal{T}(h))\} \geqslant 0. \tag{3.102}$$

证明 考虑集合 $T(h) := \mathrm{cl}\{\mathcal{T}(h)) + T_K(G(x_0))\}$. 这一集合是两个凸集合之和的拓扑包, 因而它是凸的. 进一步, 由 (3.63) 的第一包含关系, 由于外二阶切集是闭的, 有 $T(h) \subset T_K^2(G(x_0), DG(x_0)h)$. 显然, 如果在 (3.101) 中将外二阶切集用其子集 $T(h)$ 代替, 所得到最优化问题的最优值大于或等于 (3.101) 的最优值, 因此问题

$$\begin{aligned} \min_{w \in X} \quad & Df(x_0)w + D^2 f(x_0)(h,h), \\ \text{s. t.} \quad & DG(x_0)w + D^2 G(x_0)(h,h) \in T(h) \end{aligned} \tag{3.103}$$

的最优值是非负的.

最优化问题 (3.103) 是线性的, 其 (参数) 对偶 (见 (2.298)) 是

$$\max_{\lambda \in \Lambda(x_0)} \{D_{xx}^2 L(x_0, \lambda)(h,h) - \sigma(\lambda, T(h))\}. \tag{3.104}$$

事实上, (3.103) 的 Lagrange 函数是

$$\mathcal{L}(w, \lambda) = D_x L(x_0, \lambda)w + D_{xx}^2 L(x_0, \lambda)(h,h).$$

对任意的 $z \in T(h)$, 有 $z + T_K(G(x_0)) \subset T(h)$, 对任意 $\lambda \notin [T_K(G(x_0))]^- = N_K(G(x_0))$, 成立 $\sigma(\lambda, T(h)) = +\infty$. 可见 (3.103) 的参数对偶的有效域包含在 $\Lambda(x_0)$ 中, 从而得到对偶问题的形式.

进一步, Robinson 约束规范(见 (2.180)) 可推出

$$DG(x_0)X - T_K(G(x_0)) = Y.$$

由于对任何 $z \in \mathcal{T}(h), z + T_K(G(x_0)) \subset \mathcal{T}(h)$, 有

$$z + DG(x_0)X - \mathcal{T}(h) = Y,$$

因此 $DG(x_0)X - \mathcal{T}(h) = Y$. 因此, (3.103) 有可行解, 且 Robinson 约束规范对问题 (3.103) 也成立. 结果得到问题 (3.103) 与其对偶 (3.104) 间不存在对偶间隙(见定理 2.165).

我们得到 (3.104) 的最优值是非负的. 因为 $\mathcal{T}(h) \subset T(h)$, 有 $\sigma(\lambda, \mathcal{T}(h)) \leqslant \sigma(\lambda, T(h))$, 可见 (3.102) 成立, 证毕. □

如前面指出的那样 (见例 3.35), 外二阶切集 $T_K^2(G(x_0), DG(x_0)h)$ 可能是非空的. 考虑问题 (3.101) 的可行集 $[DG(x_0)]^{-1}(A)$, 其中 $A := T_K^2(G(x_0), DG(x_0)h) - D^2G(x_0)(h,h)$. 由于问题 (3.101) 的目标函数是线性的, (3.101) 的可行集可被它的凸包代替, 这不改变问题的最优值. 然而, 注意到

$$\operatorname{conv}\{[DG(x_0)]^{-1}(A)\} \subset [DG(x_0)]^{-1}[\operatorname{conv}(A)],$$

其中上述包含关系可能是严格的, 如果线性映射 $DG(x_0): X \to Y$ 不是映上的. 所以在问题 (3.101) 中, 将集合 $T_K^2(G(x_0), DG(x_0)h)$ 用它的凸包代替会改变此问题的最优值. 这是一般来讲, 在二阶条件 (3.102) 中, 当 $T_K^2(G(x_0), DG(x_0)h)$ 不是凸集时, 我们不用这一集合本身的缘故.

若 $T_K^2(G(x_0), DG(x_0)h)$ 是凸的, 则在二阶条件 (3.102) 中, 显然可以把这一集合取为 $\mathcal{T}(h)$. 在任何情况, 总可以把 $\mathcal{T}(h)$ 取成内二阶切集$T_K^{i,2}(G(x_0), DG(x_0)h)$. 我们给出注记, 若外二阶切集 $T_K^2(G(x_0), DG(x_0)h)$ 是空集, 当然集合 $\mathcal{T}(h)$ 是空集, 则对任何 $\lambda \in Y^*$ 均有 $\sigma(\lambda, \mathcal{T}(h)) = -\infty$, 因而必要条件 (3.102) 是平凡成立的. 当然, 对任何序列 $s \in \Sigma$, 可以在 (3.102) 用集合

$$\mathcal{T}^s(h) := T_K^{i,2,s}(G(x_0), DG(x_0)h).$$

由于这一集合是凸的, 则在定理 3.45 的假设下, 由 (3.102) 可推出下列二阶必要性条件

$$\inf_{s \in \Sigma} \sup_{\lambda \in \Lambda(x_0)} \{D_{xx}^2 L(x_0, \lambda)(h,h) - \sigma(\lambda, \mathcal{T}^s(h))\} \geqslant 0, \quad \forall h \in C(x_0). \tag{3.105}$$

这可推出下述结果.

命题 3.46 设点 x_0 是 (P) 的局部最优解, Robinson 约束规范在 x_0 处成立, 且 $\Lambda(x_0)=\{\lambda_0\}$. 则对所有 $h\in C(x_0)$, 成立

$$D_{xx}^2L(x_0,\lambda_0)(h,h)-\sigma(\lambda_0,T_K^2(G(x_0),DG(x_0)h))\geqslant 0. \tag{3.106}$$

证明 若 $\Lambda(x_0)=\{\lambda_0\}$, 由 (3.105) 可推出

$$\inf_{s\in\Sigma}\{D_{xx}^2L(x_0,\lambda_0)(h,h)-\sigma(\lambda_0,\mathcal{T}^s(h))\}\geqslant 0,\quad \forall h\in C(x_0).$$

由于 $\bigcup\limits_{s\in\Sigma}\mathcal{T}^s(h)=T_K^2(G(x_0),DG(x_0)h))$, 因此

$$\sup_{s\in\Sigma}\sigma(\lambda_0,\mathcal{T}^s(h))=\sigma(\lambda_0,T_K^2(G(x_0),DG(x_0)h)),$$

这证得结论. □

注 3.47 注意到, 在上述命题的条件成立时, 即使 $T_K^2(G(x_0),DG(x_0)h)$ 非凸时, 二阶必要条件 (3.106) 也是成立的.

观察到, 若

$$0\in T_K^2(G(x_0),DG(x_0)h), \tag{3.107}$$

尤其若集合 K 是多面集时, 由命题 3.34 可得

$$T_K^2(G(x_0),DG(x_0)h)=T_{T_K(G(x_0))}(DG(x_0)h).$$

进一步, 因为由临界锥的定义, 有包含关系 $DG(x_0)h\in T_K(G(x_0))$ 成立, 又由 $T_K(G(x_0))$ 是锥, 可得到 (见例 2.62)

$$T_{T_K(G(x_0))}^2(DG(x_0)h)=\mathrm{cl}\Big\{T_K(G(x_0)+[\![DG(x_0)h]\!]\Big\}.$$

在这种情形下, 取 $\mathcal{T}(h):=T_K^2(G(x_0),DG(x_0)h)$, 可得到

$$\sigma(\lambda,\mathcal{T}(h))=\begin{cases}0, & 若\ \lambda\in N_KG(x_0)\ 且\ \langle\lambda,DG(x_0)h\rangle=0,\\ +\infty, & 否则.\end{cases}$$

因为 $h\in C(x_0)$, 对任何 $\lambda\in\Lambda(x_0)$ 有 $\langle\lambda,DG(x_0)h\rangle=0$, 见命题 3.10. 所以, 此种情形, 对任何 $\lambda\in\Lambda(x_0)$ 及 $h\in C(x_0)$, 有 $\sigma(\lambda,\mathcal{T}(h))=0$, 因此 $\sigma(\lambda,\mathcal{T}(h))$ 这一项可从 (3.102) 中省略掉. 我们得到, 若 (3.107) 对每一 $h\in C(x_0)$ 均是成立的, 尤其若集合 K 是多面集, 则定理 3.45 中的二阶必要条件变为下述更熟悉的形式

$$\sup_{\lambda\in\Lambda(x_0)}D_{xx}^2L(x_0,\lambda)(h,h)\geqslant 0,\quad \forall h\in C(x_0). \tag{3.108}$$

一般地, 项 $\sigma(\lambda, \mathcal{T}(h))$ 表示为集合 K 的可能的曲率, 称为 sigma 项 (sigma term).

因为总成立 (见命题 3.34)

$$T_K^2(G(x_0), DG(x_0)h) \subset T_{T_K(G(x_0))}(DG(x_0)h),$$

由于对任何 $\lambda \in \Lambda(x_0)$ 及 $h \in C(x_0)$, 有 $\sigma(\lambda, T_{T_K(G(x_0))}(DG(x_0)h)) = 0$, 得到

$$\sigma(\lambda, \mathcal{T}(h)) \leqslant 0, \quad \text{对任意的 } \lambda \in \Lambda(x_0), h \in C(x_0). \tag{3.109}$$

所以, 一般而言, 对不一定是多面集的 K, 定理 3.45 中的必要条件要弱于 (3.108).

现在设二阶增长条件在点 $x_0 \in \Phi$ 处是成立的, 即存在 $c > 0$ 及 x_0 的邻域 N,

$$f(x) \geqslant f(x_0) + c\|x - x_0\|^2, \quad \forall x \in \Phi \cap N. \tag{3.110}$$

则在引理 3.44 的假设下, 不等式 (3.99) 可强化为

$$Df(x_0)w + D^2 f(x_0)(h, h) \geqslant \beta\|h\|^2, \tag{3.111}$$

对每一 $\beta < 2c$ 成立 ①. 此种情况下, 在定理 3.45 条件成立时, 二阶必要条件 (3.102) 可强化为

$$\sup_{\lambda \in \Lambda(x_0)} \{D_{xx}^2 L(x_0, \lambda)(h, h) - \sigma(\lambda, \mathcal{T}(h))\} \geqslant \beta\|h\|^2, \quad \forall h \in C(x_0). \tag{3.112}$$

上面的二阶条件中的 sigma 项可能是方向 h 的非连续函数. 然而, 我们有下述结果, 它在扰动分析中是有用的.

命题 3.48　设 K 是 Y 的凸子集, $\lambda \in Y^*, y \in K$. 则函数 $\psi(\cdot) := -\sigma(\lambda, T_K^{i,2}(y, \cdot))$ 是凸函数.

证明　令 $d_1, d_2 \in Y, w_1 \in T_K^{i,2}(y, d_1), w_2 \in T_K^{i,2}(y, d_2), \alpha \in [0, 1]$. 置 $d := \alpha d_1 + (1-\alpha)d_2$. 由 K 的凸性得到 $\alpha w_1 + (1-\alpha)w_2 \in T_K^{i,2}(y, d)$, 因此

$$\alpha\langle\lambda, w_1\rangle + (1-\alpha)\langle\lambda, w_2\rangle \leqslant \sigma(\lambda, T_K^{i,2}(y, d)).$$

因为 w_1, w_2 是相应的二阶切集上的任意元素, 从而

$$\alpha\sigma(\lambda, T_K^{i,2}(y, d_1)) + (1-\alpha)\sigma(\lambda, T_K^{i,2}(y, d_2)) \leqslant \sigma(\lambda, T_K^{i,2}(y, d)).$$

即函数 $\sigma(\lambda, T_K^{i,2}(y, \cdot))$ 是凹的, 因而 $\psi(\cdot)$ 是凸函数. □

对任意序列 $\sigma \in \Sigma$, 用 $T_K^{i,2,\sigma}(y, \cdot)$ 代替 $T_K^{i,2}(y, \cdot)$, 可得类似的结果.

定理 3.45 可以很容易地推广到当 X 中的闭凸集 Q 不等于 X 的情况, 可把问题 (P) 表述为下述形式

$$\text{(P)} \qquad \min_x f(x) \quad \text{s. t.} \quad (x, G(x)) \in Q \times K, \tag{3.113}$$

① 对某一 $\beta < 2c$ 即可.

此问题在可行点 x_0 处的 Lagrange 乘子集是

$$\{(\xi,\lambda):\xi+D_xL(x_0,\lambda)=0,\xi\in N_Q(x_0),\lambda\in N_K(G(x_0))\}.$$

问题 (P) 的 Lagrange 乘子集合定义为 (见 (3.16))

$$\Lambda(x_0)=\{\lambda:-D_xL(x_0,\lambda)\in N_Q(x_0),\lambda\in N_K(G(x_0))\},\tag{3.114}$$

由 (3.67) 和 (3.68) 得

$$T^{i,2}_{Q\times K}((x_0,G(x_0)),(h,DG(x_0)h))=T^{i,2}_Q(x_0,h)\times T^{i,2}_K(G(x_0),DG(x_0)h),$$

$$T^2_{Q\times K}((x_0,G(x_0)),(h,DG(x_0)h))\subset T^2_Q(x_0,h)\times T^2_K(G(x_0),DG(x_0)h).$$

由定理 3.45, 我们得到下述结果.

定理 3.49 (当 $Q\neq X$ 时的二阶必要条件的对偶形式) *设 x_0 是问题 (P) 的局部最优解 (其中 $Q\neq X$) 且 Robinson 约束规范在 x_0 处成立. 则对每一 $h\in C(x_0)$ 且对任何凸集*

$$\mathcal{T}(h)\subset T^2_{Q\times K}((x_0,G(x_0)),(h,DG(x_0)h)),$$

下述不等式成立

$$\sup_{\lambda\in\Lambda(x_0)}\{D^2_{xx}L(x_0,\lambda)(h,h)-\sigma((-D_xL(x_0,\lambda),\lambda),\mathcal{T}(h))\}\geqslant 0.\tag{3.115}$$

在 Robinson 约束规范不成立的某些情况, 仍然可以用 3.1.2 节引入的广义 Lagrange 乘子来描述二阶最优性必要条件. 这里就有这样一个结果.

定理 3.50 (没有约束规范的二阶必要性条件) *设 $\mathrm{int}(K)\neq\varnothing$. 对每一 $h\in C(x_0)$ 及任意凸集*

$$\mathcal{T}(h)\subset T^2_{Q\times K}((x_0,G(x_0)),(h,DG(x_0)h)),$$

则存在 $(\alpha,\lambda)\in\Lambda^g(x_0)$, 使得下述不等式成立

$$D^2_{xx}L^g(x_0,\alpha,\lambda)(h,h)-\sigma((-D_xL^g(x_0,\alpha,\lambda),\lambda),\mathcal{T}(h))\geqslant 0.\tag{3.116}$$

证明 设 k 是集合 K 的内部点, 考虑下述辅助的最优化问题

$$\min_{(x,t)\in Q\times\mathbb{R}} t \quad \text{s. t.}\quad f(x)-f(x_0)\leqslant t;\quad k+(1-t)(G(x)-k)\in K.$$

将此问题表示为抽象形式

$$\min_{(x,t)\in\mathcal{Q}} t \quad \text{s. t.}\quad \mathcal{G}(x,t)\in\mathcal{K},\tag{3.117}$$

其中 $\mathcal{Q} := Q \times \mathbb{R}, \mathcal{K} := K \times (-\mathbb{R}_+)$, 且

$$\mathcal{G}(x,t) := (f(x) - f(x_0) - t, k + (1-t)(G(x) - k)).$$

显然, 点 $(x_0, 0) \in X \times \mathbb{R}$ 是可行点. 进一步, 实际上 $(x_0, 0)$ 是上述问题的局部极小点. 否则会存在任意接近于 $(x_0, 0)$ 的可行点 $(\bar{x}, \bar{t})$ 满足 $\bar{t} < 0$. 点 $(\bar{x}, \bar{t})$ 的可行性可推出 $f(\bar{x}) < f(x_0)$ 且 $\bar{x} \in Q$ 及

$$G(\bar{x}) \in [k, k + (1-\bar{t})(G(\bar{x}) - k)] \subset K,$$

这可由 K 是凸集合而得到. 所以 $\bar{x}$ 是问题 (P) 的可行点, 且 $f(\bar{x}) < f(x_0)$, 这与 x_0 的局部最优性矛盾.

由于 k 是集合 K 的内部点, 有

$$\mathcal{G}(x_0, 0) + D\mathcal{G}(x_0, 0)(0, 1) = (-1, k) \in \text{int}(\mathcal{K}). \tag{3.118}$$

即 Slater 条件在 $(x_0, 0)$ 处对问题 (3.117) 的线性化成立. 问题 (3.117) 的 Lagrange 函数是

$$\mathcal{L}((x,t),(\alpha,\lambda)) = t + \alpha(f(x) - f(x_0) - t) + \langle \lambda, k + (1-t)(G(x) - k) \rangle.$$

写出相应的一阶最优条件, 可以验证 $\overline{\lambda} \in N_K(G(x_0))$, 且 $(\bar{\alpha}, \overline{\lambda}) \neq 0$, 因为

$$D_t\mathcal{L}((x_0,0),(\bar{\alpha},\overline{\lambda})) = (1 - \bar{\alpha}) + \langle \overline{\lambda}, k - G(x_0) \rangle = 0.$$

由于 $D_x\mathcal{L}((x_0,0),(\bar{\alpha},\overline{\lambda})) = 0$, 可推出 $(\bar{\alpha}, \overline{\lambda})$ 是问题 (P) 的广义 Lagrange 乘子. 再注意到, 辅助问题 (3.117) 在 $(x_0, 0)$ 的临界锥是 $C(x_0) \times \{0\}$, 其中 $C(x_0)$ 是由 (3.20) 的形式给出. 将定理 3.49 应用到问题 (3.117), 即可得到结论. □

如果 $Q = X$, 则条件 (3.116) 可表示为下述形式

$$D_{xx}^2 L^g(x_0, \alpha, \lambda)(h,h) - \sigma(\lambda, \mathcal{T}(h)) \geqslant 0, \tag{3.119}$$

其中 $\mathcal{T}(h)$ 是 $T_K^2(G(x_0), DG(x_0)h)$ 的凸子集.

3.2.3 广义的多面性

这一节讨论可以消除在定理 3.45 中给出的二阶必要条件中的 "sigma" 项的可能的情形. 如 3.2.2 节注意的那样, 若条件 (3.107) 成立, 尤其若 K 是多面集的情况则 "sigma" 项消失, 二阶必要条件是 (3.108) 的形式.

定义 3.51 *设 K 是 Banach 空间 Y 的闭凸子集. 称 K 在点 $y_0 \in K$ 处是多面性的 (polyhedric), 若对任何 $y^* \in N_K(y_0)$, 成立*

$$T_K(y_0) \cap [\ker y^*] = \text{cl}\{\mathcal{R}(y_0) \cap [\ker y^*]\}. \tag{3.120}$$

称 K 是多面性的, 若它在每一点 $y_0 \in K$ 处均是多面性的.

条件 (3.120) 意味着, 对任意切方向 $d \in T_K(y_0), \langle y^*, d\rangle = 0$, 可找到雷达方向 $d' \in \mathcal{R}(y_0)$ 满足 $\langle y^*, d'\rangle = 0$, 与 d 任意接近. 注意到, 由于 $T_K(y_0)$ 是 $\mathcal{R}(y_0)$ 的闭包, (3.120) 的左端总是包含右端的. 显然, 由定义可知, 若 $\mathcal{R}(y_0) = T_K(y_0)$, 尤其, 若 K 是多面集, 则 K 在点 y_0 处是多面性的. 本节的后面我们会看到, 在集合是非多面体的情况多面性性质也可能是成立的.

多面性的概念可以按下述方式进行拓广.

定义 3.52 考虑如下定义的具有"零曲率" (zero curvature) 的临界方向集合与雷达临界方向集 (radial critical directions)

$$C'(x_0) := \{h \in C(x_0) : 0 \in T_K^2(G(x_0), DG(x_0)h)\}, \tag{3.121}$$

$$C_{\mathcal{R}}(x_0) := \{h \in C(x_0) : DG(x_0)h \in \mathcal{R}_K G(x_0)\}. \tag{3.122}$$

称问题 (P) 满足广义的多面性条件 (强广义多面性条件) 于可行点 x_0 处成立, 若 $C'(x_0)(C_{\mathcal{R}}(x_0))$ 是 $C(x_0)$ 的稠密子集.

注意 $C_{\mathcal{R}}(x_0) \subset C'(x_0) \subset C(x_0)$, 因此, 强广义多面性条件可推出广义多面性条件. 注意到 $C_{\mathcal{R}}(x_0)$ 与 $C(x_0)$ 是凸锥. 若 $C(x_0) = \{0\}$ 或者 K 是广义的多面凸集, 则强广义多面性条件成立. 下述结果表明, 广义的多面性条件可推出 sigma 项消失.

命题 3.53 设 x_0 是问题 (P) 的局部最优解, Robinson 约束规范及广义的多面性条件在 x_0 处成立. 则下述二阶必要性条件成立

$$\sup_{\lambda \in \Lambda(x_0)} D_{xx}^2 L(x_0, \lambda)(h, h) \geqslant 0, \quad \forall h \in C(x_0). \tag{3.123}$$

证明 令 $h \in C'(x_0)$, 置 $\mathcal{T}(h) := T_K^2(G(x_0), DG(x_0)h)$. 因为 $0 \in T_K^2(G(x_0), DG(x_0)h)$, 有 $\mathcal{T}(h) = T_{T_K(G(x_0))}(DG(x_0)h))$ 是凸集, 对任何 Lagrange 乘子 λ, 有 $\sigma(\lambda, \mathcal{T}(h)) = 0$. 在 Robinson 约束规范之下, 由定理 3.45 得

$$\sup_{\lambda \in \Lambda(x_0)} D_{xx}^2 L(x_0, \lambda)(h, h) \geqslant 0, \quad \forall\, h \in C'(x_0). \tag{3.124}$$

另一方面, 函数

$$\psi(h) := \sup_{\lambda \in \Lambda(x_0)} D_{xx}^2 L(x_0, \lambda)(h, h)$$

在 X 的拓扑下是连续的. 事实上, $\psi(h)$ 是下述函数

$$\xi(h, \lambda) := D_{xx}^2 L(x_0, \lambda)(h, h)$$

在 $\lambda \in \Lambda(x_0)$ 上取上确界得到. 注意到 $\xi(h, \lambda)$ 在 $X \times Y^*$ 上是连续的, 其中 X 赋予强拓扑且 Y^* 赋予弱 * 拓扑. 进一步, 由 Robinson 约束规范得集合 $\Lambda(x_0)$ 是弱 * 紧

致的 (见定理 3.9). 于是可得 ψ 的连续性 (见 4.1 节的讨论). 由于广义多面性条件成立, $C'(x_0)$ 在 $C(x_0)$ 中是稠密的, 因而由 $\psi(\cdot)$ 的连续性可得, 对所有的 $h \in C(x_0)$, 有 $\psi(h) \geqslant 0$. □

为了讨论广义多面性与多面性的关系, 我们先考虑单位映射的情况, $G(x) := x$. 此种情形 $Y = X, y_0 = x_0$, 集合 K 是 (P) 的可行集, 且

$$C(y_0) = T_K(y_0) \cap [\ker\lambda],$$

其中 $\lambda = -Df(x_0), \lambda \in N_K(y_0)$. 显然, $C'(x_0)$ 包含 $C(x_0) \cap \mathcal{R}_K(y_0)$, 而多面性条件意味着后者是 $C(x_0)$ 的稠密子集. 因此, 对单位映射 G, K 在 $y_0 \in K$ 的多面性可推出 (P) 在 x_0 处的广义多面性质.

另一个可推出广义多面性条件的条件是

$$\operatorname{dist}(y_0 + td, K) = o(t^2), \quad \forall d \in T_K(y_0), t \geqslant 0. \tag{3.125}$$

上述条件易推出 $0 \in T_K^2(y_0, d), \forall d \in T_K(y_0)$. 显然, 即使在有限维空间的情况, 对非多面集 K, 条件 (3.125) 可能是成立的. 例如, 考虑集合

$$K := \{(y_1, y_2) \in \mathbb{R}^2 : y_2 \geqslant y_1^4\},$$

其中 $y_0 := (0, 0)$. 因此不能从 (P) 的广义多面性推出 K 的多面性条件.

并不是说 K 的多面性条件总可推出问题 (P) 的广义的多面性条件 (后面给出这样的例子). 如上面讨论, 若 G 是单位映射, 多面性可推出广义的多面性条件. 这种形式的更一般的条件在下述命题中给出.

命题 3.54 设线性映射 $DG(x_0) : X \to Y$ 是映上的, 存在 (必是唯一的) Lagrange 乘子 λ, 且 K 在 $y_0 := G(x_0)$ 处是满足多面性的. 则强广义的多面性条件在 x_0 处成立.

证明 令 $A := DG(x_0), h \in X$. 得到

$$Df(x_0)h = -\langle A^*\lambda, h\rangle = -\langle \lambda, Ah\rangle.$$

因此, $Df(x_0)h = 0$ 当且仅当 $Ah \in \ker\lambda$. 因为 $\Lambda(x_0) \neq \varnothing$, 有

$$C(x_0) = \{h \in X : Ah \in T_K(y_0), Df(x_0)h = 0\},$$

因而由 A 是映上的, 有

$$A(C(x_0)) = T_K(y_0) \cap [\ker\lambda].$$

因为 $\lambda \in N_K(y_0)$, 由多面性条件 (3.120) 得, 集合 $T_K(y_0) \cap [\ker\lambda]$ 具有稠密子集 $D \subset \mathcal{R}_K(y_0)$. 因为 A 是映上的, $C_{\mathcal{R}}(x_0) := A^{-1}(D)$ 是 $C(x_0)$ 的稠密子集①. 事实上,

① 这里的 $C_{\mathcal{R}}(x_0)$ 不是定义 3.52 中的, 它是后者的子集, 但不影响证明.

若 $C_{\mathcal{R}}(x_0)$ 在 $C(x_0)$ 中不稠密, 则存在 $C(x_0)$ 中某点的开邻域 U 满足 $U \cap C_{\mathcal{R}}(x_0)$ 是空集. 由开映射定理, 因为 A 是映上的, $A(U)$ 是 Y 的开子集, 由构造得 $A(U) \cap D = \varnothing$. 然而, 这与 D 在 $A(C(x_0))$ 中是稠密这一假设矛盾. □

下述例子对泛函空间是典型的.

例 3.55 设 Ω 是紧致的度量空间, $Y := C(\Omega)$ 是 Ω 上的连续实值函数空间, $K := C_+(\Omega)$ 是 $C(\Omega)$ 的非负值函数的集合. 现在证 K 是多面性的 (polyhedric).

令 $y \in K$. 在例 2.63 中已经证得

$$T_K(y) = \{h \in C(\Omega) : h(\omega) \geqslant 0, \forall \omega \in \Delta(y)\},$$

其中 $\Delta(y) := \{\omega \in \Omega : y(\omega) = 0\}$ 是 y 的接触点集合. 也注意到, 若 $h \in \mathcal{R}_K(y)$, 则对所有的 $\omega \in \Delta(y)$ 必有 $h(\omega) \geqslant 0$. 对任意元素 $h \in C(\Omega)$, 定义

$$h_+(\cdot) := \max\{h(\cdot), 0\}, \quad h_-(\cdot) := \min\{h(\cdot), 0\}.$$

由上述公式得到, 若 $h \in T_K(y)$, h_+ 与 h_- 均属于 $T_K(y)$. 这一性质对锥 $\mathcal{R}_K(y)$ 亦是成立的. 即若 $h \in \mathcal{R}_K(y)$, 则 h_+ 与 h_- 均属于 $\mathcal{R}_K(y)$. 实际上, 由雷达锥的定义, 有 $h \in \mathcal{R}_K(y)$ 当且仅当对充分小的 $t > 0$ 满足 $y + th \in K$. $y + th \in K$ 意味着对所有的 $\omega \in \Omega$, $y(\omega) + th(\omega) \geqslant 0$. 因为 $y(\omega) + th_+(\omega) \geqslant y(\omega) + th(\omega)$ 对所有的 $\omega \in \Omega$ 成立, 有 $h_+ \in \mathcal{R}_K(y)$. 若对某 $\omega \in \Omega$, $h(\omega) \geqslant 0$, 则 $h_-(\omega) = 0$, 从而 $y(\omega) + th_-(\omega) \geqslant 0$. 否则, $h(\omega) < 0$, 则 $h(\omega) = h_-(\omega)$, 因此还有 $y(\omega) + th_-(\omega) \geqslant 0$. 所以有 $h_- \in \mathcal{R}_K(y)$.

法锥 $N_K(y)$ 由定义在 Ω 上的非正的满足 $\operatorname{supp}(y^*) \subset \Delta(y)$ 的 Borel 测度构成. 所以, 对 $y^* \in N_K(y)$,

$$T_K(y) \cap [\ker y^*] = \{h \in Y : h(\omega) \geqslant 0, \forall \omega \in \Delta(y); h(\omega) = 0, \forall \omega \in \operatorname{supp}(y^*)\}.$$

令 $h \in T_K(y) \cap [\ker y^*]$. 由上述表达式有 h_+ 与 h_- 均属于 $T_K(y) \cap [\ker y^*]$. 进一步, 由定义, 对 $\forall \omega \in \Omega$ 有 $h_+(\omega) \geqslant 0$, 因此 $h_+ \in K$. 因为 $h \in T_K(y)$, 存在序列 $\{h_n\} \subset \mathcal{R}_K(y)$ 收敛到 h. 由上面的论证有 $(h_n)_- \in \mathcal{R}_K(y)$. 显然, 序列 $\{(h_n)_-\}$ 收敛到 h_-. 因为 $h_n \in \mathcal{R}_K(y)$, 所以对所有的 $\omega \in \Delta(y)$, 有 $h_n(\omega) \geqslant 0$, 从而对所有的 $\omega \in \Delta(y)$, 有 $(h_n)_-(\omega) = 0$ 对所有的 $\omega \in \Delta(y)$ 成立. 所以有 $(h_n)_- \in \mathcal{R}_K(y) \cap [\ker y^*]$. 最后, $\hat{h}_n := (h_n)_- + h_+$ 收敛到 h, 这一序列 (即 $\hat{h}_n$) 属于 $\mathcal{R}_K(y) \cap [\ker y^*]$. 这表明, $\mathcal{R}_K(y) \cap [\ker y^*]$ 在 $T_K(y) \cap [\ker y^*]$ 是稠密的, 因此 $C_+(\Omega)$ 是多面性的.

上述例子的论述可以正规地表述为下述抽象的形式. 令 $K \subset Y$ 是闭凸锥, 用 “$\preceq_K$” 记由 K 引导的序关系, 即 $a \preceq_K b$ 意味着 $b - a \in K$. 下述性质 (i)~(v) 对任何 $a, b, c, d \in Y$ 与 $t \in \mathbb{R}$ 成立, 若还有锥 K 是点锥(pointed)(即 $y \in K$ 且 $-y \in K$ 可推出 $y = 0$). 则性质 (vi) 亦成立:

(i) $a \preceq_K a$.

(ii) 若 $a \preceq_K b$ 且 $b \preceq_K c$, 则 $a \preceq_K c$ (传递性).

(iii) 若 $a \preceq_K b$ 且 $t \geqslant 0$, 则 $ta \preceq_K tb$.

(iv) 若 $a \preceq_K b$ 且 $t \leqslant 0$, 则 $tb \preceq_K ta$.

(v) 若 $a \preceq_K b$ 且 $c \preceq_K d$, 则 $a + c \preceq_K b + d$.

(vi) 若 $a \preceq_K b$ 且 $b \preceq_K a$, 则 $a = b$.

设锥 K 是点的. 称 $w \in Y$ 是 $a, b \in Y$ 的最小上界(或上确界), 记为 $w = a \vee b$, 若 $a \preceq_K w, b \preceq_K w$, 且对 $u \in Y$ 满足 $a \preceq_K u, b \preceq_K u$, 则 $w \preceq_K u$. 注意到, 因为 K 是点的, 由上述性质 (vi) 可得, 若最小上界 $a \vee b$ 存在, 则它是唯一的. 由定义及上述性质 (i)~(v) 得, 若对 $a, b \in Y$ 最小上界$a \vee b$ 存在, 则对 $t \geqslant 0, c \in Y$, 最小上界 $(ta) \vee (tb)$ 与 $(a+c) \vee (b+c)$ 也是存在的, 且

$$(ta) \vee (tb) = t(a \vee b), \tag{3.126}$$

$$(a+c) \vee (b+c) = (a \vee b) + c. \tag{3.127}$$

定义 3.56 令 $K \subset Y$ 是一点的闭凸锥. 称 K 在 Y 上引导格结构(lattice structure), 若对任何 $a, b \in Y$, 最小上界 $a \vee b$ 存在且算子$\vee : Y \times Y \to Y$ 是连续的.

称 $w \in Y$ 是 $a, b \in Y$ 的最大下界(或下确界), 记为 $w = a \wedge b$, 若 $w \preceq_K a, w \preceq_K b$, 且对 $u \in Y$ 满足 $u \preceq_K a, u \preceq_K b$, 则 $u \preceq_K w$. 不难证明, 若 a 与 b 的最小上界$a \vee b$ 存在, 则

$$a \wedge b = a + b - (a \vee b). \tag{3.128}$$

下确界算子$\wedge$ 满足类似 (2.126) 与 (2.127) 的性质. 记 $y_+ := y \vee 0,\ y_- := y \wedge 0$. 注意到 (3.128), 对任何 $y \in Y$,

$$y = y_- + y_+. \tag{3.129}$$

例 3.57 此例证明正半定矩阵构成的锥 $\mathcal{S}^2_+$ 不能引导在空间 $\mathcal{S}^2$ 上的格结构. 令

$$A := \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad B := \begin{bmatrix} -4 & 0 \\ 0 & 4 \end{bmatrix}, \quad W := \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix}, \quad U := \begin{bmatrix} 1 & 2 \\ 2 & 6 \end{bmatrix}.$$

则有 $A \preceq W, B \preceq W, A \preceq U, B \preceq U$, 其中"$\preceq$"记由锥 $\mathcal{S}^2_+$ 引导的 Löwner 偏序. 不难看到, 若 $V \in \mathcal{S}^2$ 满足 $A \preceq V, B \preceq V$, 且 $V \preceq W$, 则 $V = W$. 因此, 若最小上界$A \vee B$ 存在, 则它应该是 W. 然而, 矩阵 $U - W$ 不是正半定的, 即 $W \not\preceq U$. 这表明 $A \vee B$ 不存在, 因此 Löwner 偏序不具有格结构. 注意到下述定理就不会奇怪了.

定理 3.58 设 K 在 Y 上引导格结构. 则 K 是多面性的.

证明 给出如下的观察. 令 $y \in K$. 则由于 K 是锥, $\mathcal{R}_K(y) = K + [\![y]\!], N_K(y) = \{y^* \in K^- : \langle y^*, y\rangle = 0\}$.

考虑点 $h \in \mathcal{R}_K(y)$. 则 $h_+ \in \mathcal{R}_K(y), h_- \in \mathcal{R}_K(y)$. 事实上, 对 $t > 0$ 充分小, $y + th \in K$, 即 $0 \preceq_K y + th$. 因为 $h \preceq_K h_+$, 从而 $y + th \preceq_K y + th_+$, 得 $0 \preceq_K y + th_+$, 因此 $h_+ \in \mathcal{R}_K(y)$. 进一步,

$$y + th_- = y + t(h \wedge 0) = y + (th \wedge 0) = (y + th) \wedge y.$$

由于 $0 \preceq_K y + th, 0 \preceq_K y$ 得到 $0 \preceq_K (y + th) \wedge y$, 从而 $h_- \in \mathcal{R}_K(y)$.

现在设 $h \in T_K(y)$. 则 $h_+ \in T_K(y), h_- \in T_K(y)$. 事实上, 存在序列 $\{h_n\} \subset \mathcal{R}_K(y)$ 收敛于 h. 则 $(h_n)_+ \in \mathcal{R}_K(y)$, 且由算子$\vee$ 的连续性得 $(h_n)_+$ 收敛到 h_+. 因此 $h_+ \in T_K(y)$, 类似地, 对 h_- 也有类似结论.

若 $h \in \mathcal{R}_K(y), y^* \in N_K(y)$, 则 $\langle y^*, h_-\rangle = 0$. 事实上, 对充分小 $t > 0, y + th_- \in K$. 因此

$$0 \geqslant \langle y^*, y + th_-\rangle = t\langle y^*, h_-\rangle,$$

即 $\langle y^*, h_-\rangle \leqslant 0$. 另一方面, $h_- \preceq_K 0$, 从而 $-h_- \in K$. 结果有 $\langle y^*, h_-\rangle \geqslant 0$, 因此有 $\langle y^*, h_-\rangle = 0$.

现在可证 K 是多面性的. 令 $y^* \in N_K(y)$, 考虑点 $h \in T_K(y) \cap [\ker y^*]$. 则由于 $h_-, h_+ \in T_K(y)$, 有 $\langle y^*, h_-\rangle \leqslant 0, \langle y^*, h_+\rangle \leqslant 0$, 又由于

$$0 = \langle y^*, h\rangle = \langle y^*, h_+ + h_-\rangle = \langle y^*, h_+\rangle + \langle y^*, h_-\rangle,$$

得到 $h_-, h_+ \in [\ker y^*]$. 进一步, 由于 $h \in T_K(y)$, 存在序列 $\{h_n\} \subset \mathcal{R}_K(y)$ 收敛到 h. 得到 $(h_n)_- \in \mathcal{R}_K(y), (h_n)_- \in [\ker y^*]$ 且 $(h_n)_-$ 收敛到 h_-. 也有 $h_+ \in K$, 从而 $h_+ \in \mathcal{R}_K(y)$. 这样就得到序列 $\hat{h}_n := (h_n)_- + h_-$ 收敛到 h, 且这一序列属于 $\mathcal{R}_K(y) \cap [\ker y^*]$. 所以有, $\mathcal{R}_K(y) \cap [\ker y^*]$ 在集合 $T_K(y) \cap [\ker y^*]$ 中是稠密的, 因而 K 是多面性的. □

上述结果表明, 在某些泛函空间中, 非负 (非正) 值函数构成的锥是多面性的. 在 $Y := C(\Omega)$ 的情况, 我们已经看到这一点. 锥 $K := C_+(\Omega)$ 在 $C(\Omega)$ 上引导格结构,

$$(y \vee z)(\omega) = \max\{y(\omega), z(\omega)\}, \quad (y \wedge z)(\omega) = \min\{y(\omega), z(\omega)\}. \tag{3.130}$$

类似地, 在空间 $Y := L_p(\Omega, \mathcal{F}, \mu)$ 中, μ 几乎处处非负值函数构成的锥在 Y 上引导格结构, 因此是多面性的.

非负值函数锥不能引导格结构的泛函空间是连续可微函数 $\psi : \Omega \to \mathbb{R}$ 的子空间 $C^1(\Omega)$, 这是因为两个可微函数的最大值函数是不可微的.

我们将在 6.4.2 节看到上述定理应用的其他例子.

3.3 二阶充分条件

3.3.1 二阶充分性条件的一般形式

这一节设函数 $f(x)$ 与映射 $G(x)$ 是二阶连续可微的, 推导由 (3.93) 形式给出的问题 (P) 可行点的局部最优性的二阶充分条件. 我们给出论断, 这些条件也给出二阶增长条件的充分条件. 进一步, 这些充分条件是用广义 Lagrange 乘子来叙述的, 且不需要约束规范条件. 尤其不假设 Lagrange 乘子集合 $\Lambda(x_0)$ 是非空的.

为验证二阶增长条件 (3.110) 在可行点 $x_0 \in \Phi$ 处成立, 需要验证存在 $c > 0$, 对任何收敛到 x_0 的可行点序列 $x_n \in \Phi, x_n \neq x_0$, 不等式

$$f(x_n) \geqslant f(x_0) + c\|x - x_0\|^2$$

成立. 对 $\eta \geqslant 0$, 锥

$$C_\eta(x_0) := \{h \in X : \mathrm{dist}(DG(x_0)h, T_K(G(x_0))) \leqslant \eta\|h\|, Df(x_0)h \leqslant \eta\|h\|\} \quad (3.131)$$

称为近似临界锥(approximate critical cone). 对 $\eta = 0$, 上述锥与由 (3.20) 定义的临界锥 $C(x_0)$ 重合, 因此上述记号是一致的. 注意到 $C(x_0) = \bigcap\limits_{\eta>0} C_\eta(x_0)$.

下面引理表明, 可以在后续分析中假设 $x_n - x_0$ 属于近似临界锥, 当 X 是有限维时, 将 $x_n - x_0$ 表示为临界方向的扰动.

引理 3.59 设 $\{x_n\} \subset \Phi$ 是收敛到 x_0 的可行点序列且满足

$$f(x_n) \leqslant f(x_0) + o(\|x_n - x_0\|). \quad (3.132)$$

(i) 则对任何 $\eta > 0$, 对充分大的 n, 有 $x_n - x_0 \in C_\eta(x_0)$.

(ii) 若还有 X 是有限维空间, 则存在非零的临界方向 $h \in C(x_0)$ 及序列 $\{x_n\}$ 的子序列 $\{x_{n(k)}\}$, 满足 $x_{n(k)} = x_0 + t_k h + o(t_k)$ 对某一序列 $t_k \downarrow 0, t_k > 0$ 成立.

证明 由一阶 Taylor 展式, 得到

$$G(x_n) = G(x_0) + DG(x_0)(x_n - x_0) + o(\|x_n - x_0\|), \quad (3.133)$$

因为 $G(x_n) \in K, K - G(x_0) \subset T_K(G(x_0))$, 有

$$\mathrm{dist}(DG(x_0)(x_n - x_0), T_K(G(x_0))) = o(\|x_n - x_0\|).$$

进一步,

$$f(x_n) = f(x_0) + Df(x_0)(x_n - x_0) + o(\|x_n - x_0\|), \quad (3.134)$$

结合 (3.132) 可推出

$$Df(x_0)(x_n - x_0) \leqslant o(\|x_n - x_0\|).$$

从而对充分大的 n, 得到 $(x_n - x_0) \in C_\eta(x_0)$.

若还有 X 是有限维的, 则由 (i) 得 $h_n := \|x_n - x_0\|^{-1}(x_n - x_0)$ 具有极限点 h 属于 $\bigcap_{\eta>0} C_\eta(x_0) = C(x_0)$. 进一步有 $\|h\| = \lim_n \|h_n\| = 1$, 而 $t_n := \|x_n - x_0\|$ 趋于零. 这证得结论. □

由上述引理可知, 若 X 是有限维的, 可行点 x_0 满足, 相联系的临界锥为 $\{0\}$, 则 x_0 是 (P) 的局部最优解, 进一步, 一阶增长条件在 x_0 处成立. 若对 $\eta > 0$ 充分小, $C_\eta(x_0) = \{0\}$, 则对无穷维空间 X, 也有相同的结论成立. 我们已在引理 3.24 中得到这一结果.

现在回到用广义 Lagrange 函数的 Hesse 阵描述二阶充分性条件的讨论. 为讨论的方便, 引入正规化广义 Lagrange 乘子集 (the set of normalized generalized Lagrange multipliers):

$$\Lambda_N^g(x_0) := \{(\alpha, \lambda) \in \Lambda^g(x_0) : \alpha + \|\lambda\| = 1\}. \tag{3.135}$$

定义 3.60 *称二阶充分条件在可行点 $x_0 \in \Phi$ 处成立, 若存在常数 $\eta > 0$ 与 $\beta > 0$ 满足*

$$\sup_{(\alpha,\lambda)\in\Lambda_N^g(x_0)} D_{xx}^2 L^g(x_0, \alpha, \lambda)(h, h) \geqslant \beta\|h\|^2, \quad \forall h \in C_\eta(x_0). \tag{3.136}$$

称 (3.136) 的左端为广义 Lagrange 函数的极大化 (maximized)Hesse 阵. 较 (3.136) 弱的条件是

$$\forall h \in C(x_0) \setminus \{0\}, \exists (\alpha, \lambda) \in \Lambda^g(x_0) \quad \text{满足} \quad D_{xx}^2 L^g(x_0, \alpha, \lambda)(h, h) > 0. \tag{3.137}$$

因为 $\Lambda^s(x_0)$ 是锥, 奇异 Lagrange 函数$L^g(x_0, 0, \lambda)$ 关于 λ 是线性的, 这些条件较 (3.136) 要弱.

注 3.61 由于近似临界锥至少含有零方向, 二阶条件 (3.136) 隐含假设广义的 Lagrange 乘子集是非空的. 否则, (3.136) 中的上确界是 $-\infty$, 因此这些条件不可能成立. 事实上, 引理 3.24 已考虑了当近似临界锥只含有零方向时的情况, 该引理中没有假定广义 Lagrange 乘子的存在性.

注 3.62 在二阶条件 (3.136) 中, 用正规化广义 Lagrange 乘子集, 而不用广义 Lagrange 乘子集, 这一点是本质性的. 例如, 考虑一无约束问题, 令 x_0 是稳定点, 即 $Df(x_0) = 0$. 此种情况 $\Lambda^g(x_0) = \{\alpha \in \mathbb{R} : \alpha > 0\}$. 结果, 在二阶条件 (3.136) 中将 $\Lambda_N^g(x_0)$ 替换为 $\Lambda^g(x_0)$ 等价于

$$D^2 f(x_0)(h, h) > 0, \quad \forall h \in X \setminus \{0\}, \tag{3.138}$$

而条件 (3.136) 取下述形式

$$D^2 f(x_0)(h,h) \geqslant \beta\|h\|^2, \quad \forall h \in X, \tag{3.139}$$

其中 $\beta > 0$. 在 X 是无限维的情况, 二阶条件 (3.138) 对 x_0 的局部最优性不是充分的, 如例 3.66 所示. 类似地, 在例 3.25 中, 对 $x_0 = 0$, 临界锥 $C(x_0)$ 是 $\{0\}$, 因此, (3.137) 是平凡成立的. 然而, 点 $x_0 = 0$ 在此例中不是局部最优解.

下面的定理表明, (3.136) 可推出二阶增长条件, 且当 X 是有限维空间时, (3.136) 与 (3.137) 是等价的 (3.3.2 节给出另外的条件, 更适合于无穷维的最优化问题, 在此条件下 (3.136) 与 (3.137) 是等价的).

定理 3.63　(i) *设二阶充分条件 (3.136) 成立. 则二阶增长条件在点 x_0 处是成立的.*

(ii) *若 X 是有限维的, 则条件 (3.136) 与 (3.137) 是等价的.*

证明　用反证法. 设上述定理的命题 (i) 不真. 则存在序列 $x_n \in \Phi$, 具有形式 $x_n = x_0 + t_n h_n, \|h_n\| = 1, t_n \downarrow 0, t_n > 0$, 满足

$$f(x_0) + o(t_n^2) \geqslant f(x_n). \tag{3.140}$$

令 $\eta > 0$ 充分小使得 (3.136) 成立. 由引理 3.59, 对充分大的 n, 有 $h_n \in C_\eta(x_0)$, 由 (3.136), 存在有界的广义的 Lagrange 乘子 (α_n, λ_n) 满足

$$D_{xx}^2 L^g(x_0, \alpha_n, \lambda_n)(h_n, h_n) \geqslant \beta. \tag{3.141}$$

进一步, 由 $D_x L^g(x_0, \alpha_n, \lambda_n) = 0$ 及 (α_n, λ_n) 是有界的, 得到

$$L^g(x_n, \alpha_n, \lambda_n) - L^g(x_0, \alpha_n, \lambda_n) = t_n^2 D_{xx}^2 L(x_0, \alpha_n, \lambda_n)(h_n, h_n) + o(t_n^2),$$

其中 $o(t_n^2)/t_n^2 \to 0$ 相对于有界序列 (α_n, λ_n) 是一致的. 由 (3.141) 可推出

$$L^g(x_n, \alpha_n, \lambda_n) - L^g(x_0, \alpha_n, \lambda_n) \geqslant t_n^2 \beta + o(t_n^2). \tag{3.142}$$

分两种情况. 第一种情况, 若序列 $\{\alpha_n\}$ 有无穷多个非零的元素, 则若有必要, 可取子列, 有

$$f(x_n) - f(x_0) \geqslant (\alpha_n)^{-1}(L^g(x_n, \alpha_n, \lambda_n) - L^g(x_0, \alpha_n, \lambda_n)) \geqslant (\alpha_n)^{-1}(t_n^2 \beta + o(t_n^2)). \tag{3.143}$$

因为序列 $\{\alpha_n\}$ 是有界的, 与 (3.140) 矛盾. 第二种情况, 设对充分大的 n 有 $\alpha_n = 0$. 则

$$L^g(x_n, \alpha_n, \lambda_n) - L^g(x_0, \alpha_n, \lambda_n) = \langle \lambda_n, G(x_n) - G(x_0) \rangle,$$

因为 $\lambda_n \in N_K(G(x_0)), G(x_n) \in K$, 得到

$$L^g(x_n, \alpha_n, \lambda_n) - L^g(x_0, \alpha_n, \lambda_n) \leqslant 0. \tag{3.144}$$

这与 (3.142) 是矛盾的, 这就完成了 (i) 的证明.

设命题 (ii) 是不真的, 考虑函数

$$\psi(h) := \sup_{(\alpha,\lambda)\in\Lambda_N^g(x_0)} D_{xx}^2 L^g(x_0, \alpha, \lambda)(h, h).$$

则存在序列 h_n 满足 $\|h_n\| = 1, h_n \in C_{\eta(n)}(x_0)$, 其中 $\eta(n) \downarrow 0$, 且 $\lim\limits_{n\to\infty} \sup \psi(h_n) \leqslant 0$. 因为 X 是有限维的, 如果有必要, 取子列, 可设 h_n 收敛到一点 $h \in C(x_0)$. 函数 $\psi(h)$ 是一族连续函数的上确界, 因而它是下半连续的. 从而得到 $\psi(h) \leqslant 0$, 这与 (3.137) 矛盾. □

定理 3.63 的证明表明, (3.136) 形式的二阶条件成立含有两个不同的情况. 若没有 Lagrange 乘子存在, 当假设满足 (3.140) 条件的可行点序列 $x_n \to x_0$ 存在时会得到矛盾. 这得到 x_0 是满足一阶增长条件的 (P) 的局部解. 另一方面, 若存在 Lagrange 乘子, 则得到 x_0 满足二阶增长条件. 下述例子说明这一点.

例 3.64 考虑问题

$$\min_{x\in\mathbb{R}^2} \ x_1 + x_2 \quad \text{s. t.} \quad x_1^2 \leqslant 0, -x_2 \leqslant 0.$$

这是凸问题, 其可行域是 $\Phi = \{(x_1, x_2) : x_1 = 0, x_2 \geqslant 0\}$, 其最优解为 $x_0 = 0$. 这一问题的 Lagrange 乘子的集合是空集 (即所有的广义 Lagrange 乘子均是奇异的) 且 $\Lambda^g(x_0) = \{0\} \times \{(\lambda_1, \lambda_2) : \lambda_1 \geqslant 0, \lambda_2 = 0\}$. 进一步, 临界锥为

$$C(x_0) = \{(h_1, h_2) : h_1 + h_2 \leqslant 0, h_2 \geqslant 0\},$$

因为 $\Lambda_N^g(x_0) = \{0\} \times \{(1, 0)\}$, 任何非零的临界方向 h 满足 $h_1 \neq 0$, 二阶充分条件成立. 注意到, 此临界锥中的所有非零向量均是不可行的, 这与前面的讨论一致.

下述结果表明, 若 Lagrange 乘子集合 $\Lambda(x_0)$ 是非空的, 则二阶充分条件 (3.136) 可用 Lagrange 乘子来叙述.

引理 3.65 设 Lagrange 乘子的集合 $\Lambda(x_0)$ 是非空的. 则二阶充分条件 (3.136) 成立的充分必要条件是存在 $M > 0, \eta > 0, \beta > 0$($\beta$ 可能不同于 (3.136) 中的常数) 满足

$$\sup_{\|\lambda\|\leqslant M, \lambda\in\Lambda(x_0)} D_{xx}^2 L(x_0, \lambda)(h, h) \geqslant \beta\|h\|^2, \quad \forall h \in C_\eta(x_0). \tag{3.145}$$

证明　设 (3.136) 对 $\beta=\beta_0$ 时成立. 固定 $\overline{\lambda}\in\Lambda(x_0)$. 因为 $h\mapsto D^2_{xx}L(x_0,\overline{\lambda})(h,h)$ 是连续映射, 则存在 $\overline{\beta}\leqslant 0$ 满足对任何 $h\in X$, 有 $D^2_{xx}L(x_0,\lambda)(h,h)\geqslant\overline{\beta}\|h\|^2$①. 令 $t_0\in(0,1)$ 满足

$$t_0\beta_0+(1-t_0)\overline{\beta}\geqslant\frac{1}{2}\beta_0. \tag{3.146}$$

对 $(\alpha,\lambda)\in\Lambda^g_N(x_0)$, 联系着广义 Lagrange 乘子

$$(\hat{\alpha},\hat{\lambda}):=t_0(\alpha,\lambda)+(1-t_0)(1,\overline{\lambda}).$$

因为 $\hat{\alpha}=t_0\alpha+1-t_0\in(1-t_0,1)$ 是正数, Lagrange 乘子 $\tilde{\lambda}:=\hat{\alpha}^{-1}\hat{\lambda}$ 满足 $\|\tilde{\lambda}\|\leqslant M$, 其中 $M:=(1-t_0)^{-1}\max(1,\|\overline{\lambda}\|)$(注意 $\|\lambda\|\leqslant 1$). 另一方面, 由 (3.146),

$$D^2_{xx}L(x_0,\tilde{\lambda})(h,h)\geqslant\frac{1}{2}\hat{\alpha}^{-1}\beta_0\|h\|^2\geqslant\frac{1}{2}\beta_0\|h\|^2,$$

因此 (3.145) 对 $\beta:=\dfrac{1}{2}\beta_0$ 成立. 相反地, 若 (3.145) 成立, 对 $\tilde{\lambda}\in\Lambda(x_0)\cap B(0,M)$, 联系着正规化的广义 Lagrange 乘子

$$(\alpha,\lambda):=(1+\|\tilde{\lambda}\|)^{-1}(1,\tilde{\lambda}).$$

由于 $(1+\|\tilde{\lambda}\|)^{-1}\geqslant(1+M)^{-1}$, 得到

$$D^2_{xx}L^g(x_0,\alpha,\lambda)(h,h)=(1+\|\tilde{\lambda}\|)^{-1}D^2_{xx}L(x_0,\tilde{\lambda})(h,h)\geqslant(1+M)^{-1}\beta\|h\|^2,$$

取 $\beta_0:=(1+M)^{-1}\beta$, 证得 (3.136).　□

下面的例子表明, 若空间 X 是无限维的, 则二阶条件 (3.138) 不能保证 x_0 的局部最优性.

例 3.66　考虑问题

$$\min_{x\in\ell_2}\left\{f(x):=\sum_{i=1}^{\infty}(x_i^2/i-x_i^3)\right\},$$

其中 ℓ_2 是平方可积序列构成的 Hilbert 空间(如例 3.25). 则点 $x_0=0$ 满足 (3.138). 另一方面, 考虑第 k 个坐标是 $2k^{-1}$, 而其余坐标均是 0 的点列 x_k. 则 $f(x_k)=-4k^{-3}<0$, 因而 x_0 不是局部最优的.

注 3.67　在无约束情形, f 是二阶可微时, 条件 (3.139) 等价于在 x_0 点处的二阶增长条件. 考虑二次型 $Q(h):=D^2f(x_0)(h,h)$, 定义 $\|h\|_1:=Q(h)^{1/2}$. 因为双线性形式 $D^2f(x_0)(\cdot,\cdot)$ 是连续的 (由二阶可微性的定义), 因而是有界的, 则存在 $\gamma>0$ 满足 $\|h\|_1\leqslant\gamma\|h\|$. 进一步, 由 (3.139), $\|h\|_1\geqslant\beta^{\frac{1}{2}}\|h\|$. 可见 $\|\cdot\|_1$ 是 X 中与原范数

① 原著中为 $D^2_{xx}L^g(x_0,\alpha,\lambda)(h,h)$.

$\|\cdot\|$ 等价的范数. 赋予空间范数 $\|\cdot\|_1$, 空间 X 成为 Hilbert 空间, 因此 Banach 空间X 是 Hilbert 化的 (Hilbetizable), 即同胚 (homeomorphic) 于 Hilbert 空间. 我们得到这样一个惊奇的结论: 若 X 是不可 Hilbert 化的 Banach 空间, 则无约束 C^2 光滑问题在某个点 x_0 处的二阶增长条件是不可能成立的.

注 3.68 在下述两种情况下, 可分别在二阶充分条件 (3.136) 与 (3.145) 中移掉有界性条件 $|\alpha| + \|\lambda\| = 1$ 与 $\|\lambda\| \leqslant M$. 第一种情况是 X 为有限维时. 第二种情况是 Robinson 约束规范成立时, 因为此时 Lagrange 乘子集合非空且有界, 引理 3.65 可用. 再者, 因为在 Robinson 约束规范下, 可行集是度量正则的, 可以用一更小的锥来代替 $C_\eta(x_0)$,

$$\hat{C}_\eta(x_0) := \{h \in X : DG(x_0)h \in T_K(G(x_0)), Df(x_0) \leqslant \eta\|h\|\}. \tag{3.147}$$

注 3.69 条件 (3.145) 对 x_0 为下述问题的一局部最优解的充分条件:

$$\min_x \ f(x) \quad \text{s. t.} \quad G(x) \in G(x_0) + T_K(G(x_0)). \tag{3.148}$$

这是由于这样的事实: 由 (3.145) 得集合 $\Lambda(x_0)$ 是非空的凸集, 问题 (P) 与 (3.148) 有相同的 Lagrange 乘子集, 在像空间 Y 中的相应凸集的切锥也是相同的, 即在 3.2.2 节中的二阶必要条件与定理 3.63 中给出的充分条件间存在"间隙", 即 sigma 项在二阶条件 (3.136) 中不出现.

当 K 是广义多面体时, 可以消除必要与充分条件间的间隙.

定理 3.70 *设 K 是广义的多面集合, $x_0 \in \Phi$. 则*

(i) 二阶增长条件的一个充分条件是, 存在 $\beta > 0$ 满足

$$\sup_{(\alpha,\lambda)\in\Lambda_N^g(x_0)} D_{xx}^2 L^g(x_0,\alpha,\lambda)(h,h) \geqslant \beta\|h\|^2, \quad \forall h \in C(x_0). \tag{3.149}$$

(ii) 若 x_0 满足 Robinson 约束规范, 则二阶增长条件的充分必要条件是存在 $\beta > 0$,

$$\sup_{\lambda\in\Lambda(x_0)} D_{xx}^2 L(x_0,\lambda)(h,h) \geqslant \beta\|h\|^2, \quad \forall h \in C(x_0). \tag{3.150}$$

证明 (i) 根据定理 3.63, 只需证若 K 是广义多面体, 则 (3.149) 可推出 (3.136). 这是一个将定理 2.200 用于临界锥可很容易得到的结论. 因为给定 $h \in C_\eta(x_0)$, 存在 $\hat{h} \in C(x_0)$ 满足 $\|\hat{h} - h\| = O(\eta)$. 用到可轻易证明的结论, 即由 (3.149) 中的上确界函数在 X 中的单位球内是一致连续的, 可证得 (i).

(ii) 由定理 3.45, 因为 K 是广义的多面集, 所以 sigma 项是零, $\beta = 0$ 时 (3.150) 是二阶必要条件. 由于引理 3.44 的证明中构造的路径 $x(t)$ 满足 $\|x(t) - x_0\| =$

$t\|h\|+o(t)$, 得到二阶增长条件的必要性条件是：存在不依赖于 h 的某一 $\beta>0$, (3.103) 的值大于或等于 $\beta\|h\|^2$. 由定理 3.45 的证明中对偶性的论述, 得到 (3.150) 是二阶增长条件的必要条件.

此条件的充分性可由 (i) 结合引理 3.65 得到 (在 $\eta=0$ 时证明不发生变化), 注意到 Robinson 约束规范可推出 Lagrange 乘子集合是有界的, 因而在 (3.145)(其中取 $\eta=0$) 中对充分大的常数 M, 得到 (3.150). □

除了 K 是广义多面体的情况, 还有两个基本的情况可能消除二阶必要性条件与二阶充分性条件间的间隙：一种情况是 x_0 处广义多面性质成立的情形, 第二种情况是当 X 是有限维空间且集合 K 满足我们称为二阶正则性的性质 (second order regularity). 在 3.3.3 节中将讨论这种情形.

3.3.2　二次的 Legendre 形式与广义的 Legendre 形式

这一节讨论可消除 (3.136) 形式的二阶充分条件与较弱形式的 (3.137) 间的间隙的某些条件. 函数 $B:X\times X\to\mathbb{R}$ 称为双线性的, 若对任何 $x\in X$, 函数 $B(\cdot,x)$ 与 $B(x,\cdot)$ 在 X 上是线性的. 双线性函数 B 称为对称的 (symmetric), 若对任何 $x_1,x_2\in X$ 有 $B(x_1,x_2)=B(x_2,x_1)$. 称函数 $Q:X\to\mathbb{R}$ 为 X 上的二阶形式, 若存在双线性的对称函数 $B(\cdot,\cdot)$ 满足 $Q(x)=B(x,x)$. 若对所有的 $x\in X$, 有 $Q(x)\geqslant 0$, 则二次形式是非负的. 对所有的 $x\in X\setminus\{0\}$, 有 $Q(x)>0$, 则称二次形式 Q 是正的. 用

$$\mathcal{N}(Q):=\{x\in X:Q(x)=0\}$$

记二次形式Q 的零空间.

命题 3.71　*二次形式 $Q(\cdot)$ 在 X 上是凸的当且仅当它是非负的.*

证明　设 B 是对应于 Q 的双线性函数. 有 Q 是凸的当且仅当对任意 $x,h\in X$, 函数 $\phi(t):=Q(x+th)$ 在 $\mathbb{R}$ 上是凸的. 可将 ϕ 表示为

$$\phi(t)=t^2Q(h)+2tB(x,h)+Q(x).$$

即 ϕ 是二次形式, 它是凸的当且仅当 $Q(h)\geqslant 0$, 这证得结论. □

由上述命题得, 若非负的二次型Q 是下半连续的, 因而其上图是凸的且以强拓扑为闭的, 则其上图以弱拓扑为闭的, Q 是弱下半连续的. 由命题 2.111 还可得到, 若 Q 是一非负的下半连续的二次型式, 则 Q 是连续的.

命题 3.72　*若 Q 是非负的二次形式, 则 $\mathcal{N}(Q)$ 是 X 的线性子空间, 且对任何 $h\in\mathcal{N}(Q)$ 与 $x\in X$, 有 $Q(x+h)=Q(x)$.*

证明　若 $h\in\mathcal{N}(Q),t\in\mathbb{R}$, 则 $Q(th)=t^2Q(h)=0$, 因此 $th\in\mathcal{N}(Q)$. 再注意到 Q 是非负的凸的, 因此 $\mathcal{N}(Q)$ 是凸锥. 有若 $h\in\mathcal{N}(Q)$, 则 $-h\in\mathcal{N}(Q)$. 得到 $\mathcal{N}(Q)$ 是线性空间.

设 B 是对应 Q 的双线性函数, $h \in \mathcal{N}(Q)$. 因为 Q 是非负的凸的, $Q(h)=0$, 则对任何 $t \in [0,1], x \in X$, 有

$$tQ(x)+(1-t)Q(h) \geqslant Q(tx+(1-t)h)=t^2Q(x)+2t(1-t)B(x,h).$$

得到 $Q(x) \geqslant 2B(x,h)$. 由于 $\mathcal{N}(Q)$ 是线性空间, $B(x,\cdot)$ 是线性的, 只有在 $B(x,h)=0$ 时此不等式才能成立. 再注意到 $Q(x+h)=Q(x)+Q(h)+2B(x,h)$, 得 $Q(x+h)=Q(x)$. □

显然, 若 Q 是连续的或非负的且 l.s.c. 的, 则集合 $\mathcal{N}(Q)$ 是闭的.

定义 3.73 *称函数 $Q: X \to \mathbb{R}$ 是广义的 Legendre 形式, 若它是弱下半连续 2 度正齐次的, 即 $\forall x \in X, t>0, Q(tx)=t^2Q(x)$, 且若 $x_k \xrightarrow{w} x$ 及 $Q(x_k) \to Q(x)$, 则 $x_k \to x$.*

注意到二次形式总是 2 度正齐次的. 若 Q 是二次形式, 上述定义即简化为 Legendre 形式, 它是变分计算理论中一个经典的研究对象. 在第 6 章, 当讨论最优控制问题时, 我们再来说明本节中一些稍抽象的结论.

命题 3.74 *设 X 是自反的Banach 空间. 设在 (3.136) 中的广义 Lagrange 函数的极大化 Hesse 阵*

$$Q(\cdot) := \sup_{(\alpha,\lambda)\in\Lambda_N^g(x_0)} D_{xx}^2 L^g(x_0,\alpha,\lambda)(\cdot,\cdot) \tag{3.151}$$

是广义的Legendre 形式. 则条件 (3.136) 与 (3.137) 是等价的.

上述命题可立即由下述引理得到, 而此引理本身也是颇有用途的.

引理 3.75 *设 X 是自反的Banach 空间, $Q: X \to \mathbb{R}$ 是广义的 Legendre 形式, $C \subset X$ 是凸闭锥, C_η , $\eta>0$ 是满足下述性质的一族锥: 若对某序列 $\eta_k \downarrow 0$, $h_k \in C_{\eta_k}$, 序列 $\{h_k\}$ 弱收敛到 $\bar{h}$, 则 $\bar{h} \in C$. 设对所有的 $h \in C\setminus\{0\}$, 有 $Q(h)>0$. 则存在正的常数 β 与 $\bar{\eta}$, 满足对任意 $h \in C_{\bar{\eta}}$, 有 $Q(h) \geqslant \beta\|h\|^2$.*

证明 假设结论不真. 令 η_k 是收敛于 0 的正数序列, 令 $C_k := C_{\eta_k}$. 则存在序列 $h_k \in C_k, h_k \neq 0$ 满足 $Q(h_k) \leqslant k^{-1}\|h_k\|^2$, 对 k 充分大成立. 因为 Q 是 2 度正齐次的, 可设 $\|h_k\|=1$. 由于 X 是自反的, 序列 h_k 至少有一弱极限点$\bar{h}$. 如有必要, 可取一子序列, 可设 h_k 弱收敛到 $\bar{h}$. 锥族 C_η 的性质可推出 $\bar{h} \in C$. 由 Q 的弱下半连续性, 有

$$Q(\bar{h}) \leqslant \liminf_{k\to\infty} Q(h_k) \leqslant 0. \tag{3.152}$$

若 $\bar{h}=0$, 则 $Q(\bar{h})=0$, 由 (3.152) 可得 $Q(h_k) \to 0$. 由定义 3.7.3, 这导致 $h_k \to 0$ 与 $\|h_k\|=1$ 矛盾. 因此 $\bar{h} \neq 0$ 且 $h \in C$, 从而由假设有 $Q(\bar{h})>0$. 这显然与 (3.152) 矛盾, 这就完成了证明. □

现在给出刻画广义 Legendre 形式的准则. 称二次型 Q 是椭圆型的, 若 Q 是连续的, 且存在 $\alpha > 0$, 满足

$$Q(x) \geqslant \alpha \|x\|^2, \quad \forall x \in X.$$

显然椭圆型二次形式是正的, 因而是凸的, 结果在弱拓扑下是下半连续的. 注意到, 若 Q 是椭圆型二次形式, 则 $\|\cdot\|_1 := Q(\cdot)^{1/2}$ 在空间 X 上定义了范数, 满足 $\|\cdot\|_1 \geqslant \alpha^{1/2}\|\cdot\|$. 因为 Q 是连续的, 有 $\|\cdot\|_1 \leqslant \gamma\|\cdot\|$, 其中 γ^2 是与 Q 联系的双线性形式的范数. 得到两个范数 $\|\cdot\|$ 与 $\|\cdot\|_1$ 是等价的, 而赋予范数 $\|\cdot\|_1$, X 成为 Hilbert 空间. 所以由椭圆型二次形式的存在可推出空间 X 是可 Hilbert 化的 (Hilbertizable).

命题 3.76 (i) 任何椭圆型的二次形式是 Legendre 形式.

(ii) 令 Q_1 是广义的 Legendre 形式, Q_2 是弱连续的且是 2 度正齐次的. 则 $Q := Q_1 + Q_2$ 是广义的 Legendre 形式.

证明 (i) 设 Q 是椭圆型的二次形式, 则 Q 是凸的实值的, 在弱拓扑下是下半连续的. 令 $B(\cdot,\cdot)$ 是与 Q 相联系的双线性形式. 由于

$$B(x_1, x_2) = \frac{1}{4}[Q(x_1 + x_2) - Q(x_1 - x_2)],$$

双线性形式 B 也是连续的. 令 $\{x_k\}$ 是 X 中的序列, 满足 $x_k \xrightarrow{w} x$ 且 $Q(x_k) \to Q(x)$. 有

$$\alpha\|x_k - x\|^2 \leqslant Q(x_k - x) = Q(x_k) - 2B(x_k, x) + Q(x).$$

由于 $B(\cdot, x)$ 在 X 上是连续的线性泛函, 因此, 由 $x_k \xrightarrow{w} x$ 得 $B(x_k, x) \to B(x, x) = Q(x)$. 这就得到 $\|x_k - x\| \to 0$, 证得 (i).

(ii) 若 Q_1 是广义的 Legendre 形式, 它是弱下半连续的且 2 度正齐次的. 结果, 由于 Q_2 是弱连续的且 2 度正齐次的, 我们得到和函数 Q 是弱下半连续的且 2 度正齐次的. 令 $x_k \xrightarrow{w} x, Q(x_k) \to Q(x)$. 由于 Q_2 是弱连续的, 有 $Q_2(x_k) \to Q_2(x)$. 因此, $Q_1(x_k) \to Q_1(x)$. 再由 Q_1 是广义的 Legendre 形式, 得 $x_k \to x$. □

命题 3.77 设

(i) Hesse 阵 $D^2 f(x_0)(\cdot,\cdot)$ 是 Legendre 形式.

(ii) 在 X 的弱拓扑与 Y 的强拓扑下, 映射 $h \mapsto D_{xx}^2 G(x_0)(h, h) : X \to Y$ 是连续的.

则对满足集合 $\{\lambda \in \Lambda(x_0) : \|\lambda\| \leqslant M\}$ 非空的任何 $M > 0$,

$$Q(h) := \sup_{\substack{\|\lambda\| \leqslant M \\ \lambda \in \Lambda(x_0)}} D_{xx}^2 L(x_0, \lambda)(h, h)$$

是广义的 Legendre 形式.

证明 考虑定义在 $Y^* \times X$ 上的实值函数 $\psi(\lambda, h) := \langle \lambda, D^2_{xx}G(x_0)(h,h)\rangle$. 由假设 (ii), $\psi(\lambda, h)$ 在以 Y^* 的弱 * 拓扑与 X 的弱拓扑的积上连续. 集合 $C := \{\lambda \in \Lambda(x_0) : \|\lambda\| \leqslant M\}$ 是弱 * 紧致的. 结果, 函数 $v(h) := \sup_{\lambda \in C} \psi(\lambda, h)$ 以 X 的弱拓扑连续 (见 4.1 节的命题 4.4). 显然 $v(h)$ 是 2 度正齐次的. 因为 Q 是 Legendre 形式 $D^2_{xx}f(x_0)(h,h)$ 与 $v(h)$ 的和, 由命题 3.76(ii) 得 Q 是广义 Legendre 形式. □

注 3.78 在命题 3.77 的假设下, 由命题 3.74 可得, 条件 (3.136) 与 (3.137)(应用于 Lagrange 函数而不是广义的 Lagrange 函数) 是等价的.

现在给出 Legendre 形式的一些额外刻画. 称二次形式$Q : X \to \mathbb{R}$ 具有有限的秩 n(finite rank n), 若存在二次形式 $Q_1 : \mathbb{R}^n \to \mathbb{R}$ 与线性的连续算子$A : X \to \mathbb{R}^n$, 满足对所有的 $x \in X$ 有 $Q(x) = Q_1(Ax)$. Hilbert 空间X 的线性子空间 W 有有限的余维数 n, 若存在维数是 n 的线性空间 $V \subset X$ 满足 $W = V^\perp$. 注意, 由此定义, 具有有限余维数的 X 的子空间是闭的.

命题 3.79 *设 X 是 Hilbert 空间, $Q : X \to \mathbb{R}$ 是二次形式, 则下述条件是等价的:*

(i) 二阶形式 Q 是 Legendre 形式.

(ii) Q 到 X 中的任何闭子空间上的限制是 Legendre 形式.

(iii) Q 到 X 中的某有限余维数的子空间的限制是椭圆型的二次形式.

(iv) 二次形式 Q 是椭圆型二次形式与具有有限秩的二次形式的和.

证明 条件 (i) 与 (ii) 的等价性可由 Legendre 形式的定义立即得到. 现在设 (ii) 成立. 为证明 (ii)⇒(iii), 需要证明存在有限余维数的 X 的子空间 W, 存在正常数 α 满足 $Q(x) \geqslant \alpha\|x\|^2, \forall x \in W$. 事实上, 此种情形, Q 在 W 上的限制是凸的, 由 Q 是弱下半连续的, 这一限制是连续的, 因而是椭圆型的.

考虑问题

$$\min_{x \in X} Q(x) \quad \text{s. t.} \quad \|x\| = 1. \tag{3.153}$$

这一问题有正的最优值 α 当且仅当对所有的 $x \in X$, 有 $Q(x) \geqslant \alpha\|x\|^2$. 此种情形, Q 是椭圆型的, 即证得结论. 假设最优值是非正的. 令 $\{x_k\}$ 是上述问题的极小化序列, 即 $\|x_k\| = 1, k = 1, 2, \cdots$, 且 $Q(x_k)$ 收敛到上述问题的最优值. 因为 $\{x_k\}$ 是有界的, 它至少有弱极限点$\bar{x}$. 我们断定 $Q(\bar{x}) \leqslant 0$ 且 $\bar{x} \neq 0$. 事实上, 因为 Q 是 Legendre 形式, 因而是弱下半连续的, 得到 $Q(\bar{x}) \leqslant \lim_{k\to\infty} Q(x_k) \leqslant 0$. 若 $Q(\bar{x}) < 0$, 则由于 $Q(0) = 0$, 得到 $\bar{x} \neq 0$. 若 $Q(\bar{x}) = 0$, 则得到 $Q(x_k) \to Q(\bar{x})$. 此种情形, 由于 Q 是 Legendre 形式, 得到 $x_k \to \bar{x}$, 从而 $\|\bar{x}\| = 1$, 因此 $\bar{x} \neq 0$.

现在考虑下述迭代过程. 置 $x_1 := \bar{x}/\|\bar{x}\|$, 令 W_1 是 X 中与 x_1 垂直的子空间. 若 Q 到 W_1 上的限制是椭圆型的, 就证得结论. 否则, 类似于上面的做法, 存在 $x_2 \in W_1$ 满足 $\|x_2\| = 1, Q(x_2) \leqslant 0$. 考虑与 x_1 和 x_2 垂直的空间 W_2 等. 此过程或者经过

有限步终止, 此时证得结论, 或者得到序列 $\{x_k\}$, 满足 $\|x_k\| = 1, k = 1, 2, \cdots$, 且 $\langle x_k, x_m\rangle = 0, \forall k \neq m$. 则有 $x_k \xrightarrow{w} 0$. 事实上, 对任何 $m \in \mathbb{N}, \lim_k \langle x_k, x_m\rangle = 0$, 因此对由向量 $x_1, x_2, \cdots$ 生成的线性空间 V 中所有的 x 有 $\lim_k\langle x, x_k\rangle = 0$. 因为 $\{x_k\}$ 是有界的, 则对所有的 $x \in \mathrm{cl}\ V$, 有 $\lim_k\langle x, x_k\rangle = 0$. 若 $x \in V^\perp$, 则显然 $\lim_k\langle x, x_k\rangle = 0$, 这得到 $x_k \xrightarrow{w} 0$. 由于 $x_k \xrightarrow{w} 0$ 及 Q 是弱下半连续的, 得 $0 = Q(0) \leqslant \liminf\limits_{k\to\infty} Q(x_k) \leqslant 0$. 结果, $Q(x_k) \to Q(0)$, 因此由于 Q 是 Legendre 形式, 得 $x_k \to 0$, 矛盾. 这完成了 (ii)⇒(iii) 的证明.

现在设 (iii) 成立. 则可将 X 表示为 $X = X_1 \oplus X_2$, 其中 Q 在 X_1 上是椭圆型的, X_2 是有限维的. 将 $x \in X$ 分解为 $x = x_1 + x_2, x_1 \in X_1, x_2 \in X_2$. 记 $a(\cdot,\cdot)$ 为与 Q 相联系的双线性形式. 置 $Q_1(x) := Q(x_1) + \|x_2\|^2, Q_2(x) := Q(x_2) - \|x_2\|^2 + a(x_1, x_2)$. 则 $Q(x) = Q_1(x) + Q_2(x)$ 且 Q_1 是椭圆型的. 以下验证 $Q_2(x)$ 是有限秩的. $a(x_1, \cdot)$ 是 X_2 上的线性的连续的泛函, 存在 $x_2^* \in X_2$ 满足 $a(x_1, \cdot) = \langle \cdot, x_2^*\rangle$. 容易验证映射 $A_1x_1 := x_2^*$ 是线性的且连续的. 因此有 $a(x_1, x_2) = \langle x_2, A_1x_1\rangle$, 其中 $A_1 : X_1 \to X_2$ 是线性的连续的算子. 置 $Ax := (A_1x_1, x_2) : X \to X_2 \times X_2$. 显然, $X_2 \times X_2$ 是有限维的, 且 $Q_2(x) = \widehat{Q_2}(Ax)$, 其中 $\widehat{Q_2}(y, z) := Q_2(z) - \|z\|^2 + \langle y, z\rangle$. 因此, 由 (iii) 可推出 (iv).

最后, 推出关系 (iv)⇒(i) 可由命题 3.76 证得. □

命题 3.80 是上述命题证明的一个结果.

命题 3.80 *设 X 是 Hilbert 空间, $Q : X \to \mathbb{R}$ 是非负的二阶形式. 则 Q 是 Legendre 形式当且仅当 Q 的零空间 $\mathcal{N}(Q)$ 是有限维的, Q 在 $\mathcal{N}(Q)$ 的正交补空间上是椭圆型的.*

注 3.81 设 X 是 Hilbert 空间, $Q : X \to \mathbb{R}$ 是 Legendre 形式, A 是相应的自伴随 (self adjoint) 连续的线性算子, 定义为 $\langle Ax, x'\rangle = B(x, x')$, 其中 B 是与 Q 相联系的双线性函数. 若 Q 不是椭圆型的, 则像命题 3.79 的证明那样, 存在问题 (3.153) 的最优解 x_1, 满足 $\|x_1\| = 1$ 且 $Q(x_1) \leqslant 0$. 由一阶最优性条件得 x_1 是 A 的特征向量, 即 $Ax_1 = \lambda_1 x_1$, 其中 $\lambda_1 = Q(x_1)$. 得到 $A(W_1) \subset W_1$, 其中 W_1 是与 x_1 垂直的线性空间. 由归纳得下述的分解

$$Q(x) = \sum_{i=1}^{l} \lambda_i \langle x_i, x\rangle^2 + Q_2(\pi(x)),$$

其中 $\lambda_i \leqslant 0, i = 1, \cdots, l, \pi(x)$ 是 x 到空间 $\{x_1, \cdots, x_l\}^\perp$ 上的直交投影, Q_2 是 $\{x_1, \cdots, x_l\}^\perp$ 上的椭圆型形式.

3.3.3 集合的二阶正则性与"无隙"二阶最优性条件

这一节我们讨论一种情形, 此种情形可弥补 (3.102) 形式的二阶必要条件与二

阶充分条件间的间隙. 二阶必要性条件 (3.102) 基于目标函数沿 (3.96) 形式的可行的抛物路径上的上方估计. 为了得到充分条件, 我们需要导出目标函数的下估计. 如只考虑抛物路径的下估计, 这在事先是没有根据的. 然而, 对一类重要问题, 包括半定与一大类半无限规划问题在内, 这种做法是可以的. 尽管二阶正则性的概念是在 Banach 空间的框架下引入的, 我们得到当 X 是有限维的情况下所期望的下估计.

定义 3.82 设 S 是 Banach 空间 Y 中的闭子集, $y \in S, d \in T_S(y)$, 考虑连续的线性映射 $M : X \to Y$. 称闭集合$\mathcal{A}_{S,M}(y,d) \subset Y$ 是 S 在点 y 处沿方向 d 关于 M 的上二阶近似集 (upper second order approximation set), 若对具有形式 $y_n := y + t_n d + \dfrac{1}{2}t_n^2 r_n$ 的任何序列 $y_n \in S$, 其中 $t_n \downarrow 0, r_n = Mw_n + a_n, \{a_n\}$ 是 Y 中的收敛序列, $\{w_n\}$ 是 X 中的满足 $t_n w_n \to 0$ 的序列, 下述条件成立

$$\lim_{n\to\infty} \operatorname{dist}(r_n, \mathcal{A}_{S,M}(y,d)) = 0. \tag{3.154}$$

若上述等式对 $Y = X$ 且 M 是单位映射成立, 即对任何满足 $t_n r_n \to 0$ 的序列 $y + t_n d + \dfrac{1}{2}t_n^2 r_n \in S$ 有条件 (3.154) 成立, 略掉 M, 称集合 $\mathcal{A}_S(y,d)$ 是集合 S 在点 y 处沿方向 d 的上二阶近似.

上述定义旨在构造充分大集合 $\mathcal{A}_S(y,d)$ 满足若 $y + td + \varepsilon(t)$ 是 S 中与 d 相切的曲线, 其中 $\varepsilon(t) = o(t)$, 则二阶余项 $r(t) := \left(\dfrac{1}{2}t^2\right)^{-1}\varepsilon(t)$ 当 $t \downarrow 0$ 时趋于 $\mathcal{A}_S(y,d)$. 注意到余项 $r(t)$ 与其序列形式 $r_n := r(t_n)$ 可能是无界的. 在无穷维空间, 由于技术上的原因, 需要考虑线性映射 M 等额外的复杂性.

做如下的观察. 上二阶近似集 $\mathcal{A}_S(y,d)$ 是不唯一的. 显然, 若 $\mathcal{A}_S(y,d) \subset B$, 则 B 也是上二阶近似集合. 因为若 $y \in S, d \in T_S(y), y{+}d{+}w \in S$ 可推出 $d{+}w \in T_S(y)$, 因此有 $w \in T_{T_S(y)}(d)$, 这得到集合 $T_{T_S(y)}(d)$ 总是上二阶近似集合. 不难由定义看到, 外二阶切集 $T_S^2(y,d)$ 包含在任何上二阶近似集 $\mathcal{A}_S(y,d)$ 内.

定理 3.83 设空间 X 是有限维的. 令 x_0 是问题 (P) 的可行点满足广义的 Lagrange 乘子集合 $\Lambda^g(x_0)$ 是非空的. 令每一 $h \in C(x_0)$ 对应 K 在点 $y_0 := G(x_0)$ 处的沿方向 $d := DG(x_0)h$ 关于线性映射 $M := DG(x_0)$ 的上二阶近似集 $\mathcal{A}(h) := \mathcal{A}_{K,M}(y_0,d)$, 设下述二阶条件成立: 对任何 $h \in C(x_0) \setminus \{0\}$, 存在 $(\alpha,\lambda) \in \Lambda^g(x_0)$, 满足

$$D_{xx}^2 L^g(x_0,\alpha,\lambda)(h,h) - \sigma(\lambda, \mathcal{A}(h)) > 0. \tag{3.155}$$

则在 x_0 处的二阶增长条件成立, 因而 x_0 是问题 (P) 的严格局部最优解.

证明 用反证法. 假设二阶增长条件在 x_0 处不成立. 则存在可行点序列 $x_n \in \Phi$ 收敛到 x_0, 满足

$$f(x_n) \leqslant f(x_0) + o(t_n^2), \tag{3.156}$$

其中 $t_n := \|x_n - x_0\|$. 由近似临界锥条件的紧致性, 若有必要可取一子列, 可设 $h_n := (x_n - x_0)/t_n$ 收敛到向量 $h \in C(x_0)$(见引理 3.59 的证明). 显然, $\|h\| = 1$, 因而 $h \neq 0$.

由 $G(x_n)$ 在 x_0 处二阶 Taylor 展式, 有

$$G(x_n) = y_0 + t_n d + \frac{1}{2}t_n^2(DG(x_0)w_n + D^2G(x_0)(h,h)) + o(t_n^2),$$

其中 $w_n := 2t_n^{-2}(x_n - x_0 - t_n h)$. 注意到 $x_n - x_0 - t_n h = o(t_n)$, 有 $t_n w_n \to 0$.

结合上二阶近似集合的定义, 可以得到

$$DG(x_0)w_n + D^2G(x_0)(h,h) \in \mathcal{A}(h) + o(1)B_Y. \tag{3.157}$$

也有

$$f(x_n) = f(x_0) + t_n Df(x_0)h + \frac{1}{2}t_n^2(Df(x_0)w_n + D^2f(x_0)(h,h)) + o(t_n^2),$$

因此, 用 (3.156) 与 (3.157), 可找到序列 $\varepsilon_n \to 0$ 满足

$$\begin{cases} 2t_n^{-1}Df(x_0)h + (Df(x_0)w_n + D^2f(x_0)(h,h)) \leqslant \varepsilon_n, \\ DG(x_0)w_n + D^2G(x_0)(h,h) \in \mathcal{A}(h) + \varepsilon_n B_Y. \end{cases} \tag{3.158}$$

由 (3.155), 存在 $(\alpha, \lambda) \in \Lambda^g(x_0)$ 满足

$$D_{xx}^2 L^g(x_0, \alpha, \lambda)(h,h) - \sigma(\lambda, \mathcal{A}(h)) \geqslant \kappa, \tag{3.159}$$

其中 $\kappa > 0$ 是数. 由 (3.158) 的第二条件得

$$\langle \lambda, DG(x_0)w_n + D^2G(x_0)(h,h)\rangle \leqslant \sigma(\lambda, \mathcal{A}(h) + \varepsilon_n B_Y) = \sigma(\lambda, \mathcal{A}(h)) + \varepsilon_n\|\lambda\|.$$

有 $\alpha \geqslant 0$, 若 $\alpha \neq 0$, 则存在 Lagrange 乘子, 因此有 $Df(x_0)h = 0$. 任何情况都有 $\alpha Df(x_0)h = 0$, 因此, 由 (3.158) 与 (3.159) 可得

$$\begin{aligned} 0 \geqslant& \alpha(2t_n^{-1}Df(x_0)h + Df(x_0)w_n + D^2f(x_0)(h,h) - \varepsilon_n) \\ &+ \langle \lambda, DG(x_0)w_n + D^2G(x_0)(h,h)\rangle - \sigma(\lambda, \mathcal{A}(h)) - \varepsilon_n(\|\lambda\|) \\ =& D_{xx}^2 L^g(x_0, \alpha, \lambda)(h,h) - \sigma(\lambda, \mathcal{A}(h)) - \varepsilon_n(\alpha + \|\lambda\|) \\ \geqslant& \kappa - \varepsilon_n(\alpha + \|\lambda\|). \end{aligned}$$

因为 $\varepsilon_n \to 0$, 得到矛盾, 这就完成了证明. □

若 Lagrange 乘子集合 $\Lambda(x_0)$ 是非空的, 则二阶充分条件 (3.155) 等价于

$$\sup_{\lambda \in \Lambda(x_0)} \{D_{xx}^2 L(x_0, \lambda)(h,h) - \sigma(\lambda, \mathcal{A}(h))\} > 0, \quad \forall h \in C(x_0) \setminus \{0\}. \tag{3.160}$$

如前面指出的那样, 集合 $Z(h) := T_{T_K(G(x_0))}(DG(x_0)h)$ 总是上二阶近似集. 进一步,

$$\sigma(\lambda, Z(h)) = \begin{cases} 0, & 若\ \lambda \in N_K(G(x_0)), \langle \lambda, DG(x_0)h \rangle = 0, \\ +\infty, & 否则. \end{cases}$$

所以, 对上二阶近似集的这一种选取, 二阶充分条件 (3.155) 与二阶条件 (3.137) 是相同的.

例 3.84 考虑在例 3.29 构造的集合 S 与最优化问题

$$\min f(x) \quad \text{s. t.} \quad x \in S,$$

其中 $f : \mathbb{R}^2 \to \mathbb{R}$ 是二次连续可微函数. 设 $x_0 := (0, 0)$ 是上述问题的稳定点, 即 $-\nabla f(x_0) \in N_S(x_0)$. 注意到 $\lambda := -\nabla f(x_0)$ 是 (唯一的)Lagrange 乘子, $N_S(x_0) = \{(x_1, x_2) : x_1 = 0, x_2 \leqslant 0\}$. 进一步, 设 $\nabla f(x_0) \neq 0$. 则临界锥 $C(x_0)$ 等于 $\{(x_1, x_2) : x_2 = 0\}$. 对任何 $h \in C(x_0)$ 与 $a \in \mathbb{R}$, 集合 $\mathcal{A}(h) := \{(x_1, x_2) : x_2 \geqslant a\}$ 是 S 在 x_0 处沿 h 的外二阶近似集. 对任何给定的 $h \in C(x_0)$, sigma 项 $\sigma(\lambda, \mathcal{A}(h))$ 可以任意地小, 因此, 当 a 充分大时 $-\sigma(\lambda, \mathcal{A}(h))$ 可以任意大. 结果, 无论 $\nabla^2 f(x_0)$ 的值是多少, 充分条件 (3.155) 成立. 于是得到, 不管 $\nabla^2 f(x_0)$ 的值是多少, 稳定点 x_0 是上述问题的局部最优解. 在此例中, 对任何 $h \in C(x_0) \setminus \{0\}$, 二阶切集 $T_S^{i,2}(x_0, h)$ 与 $T_S^2(x_0, h)$ 均是空集, 因此二阶必要条件 (3.102) 是成立的.

注意, 此例中, 若 $\nabla f(x_0) = 0$, 则 $\lambda = 0$, 对任何上二阶近似集 $\mathcal{A}(h)$, sigma 项 $\sigma(\lambda, \mathcal{A}(h))$ 是 0.

比较分别由 (3.102) 与 (3.160) 给出的二阶必要条件与充分条件, 我们可以观察到, 除了由弱不等式到严格不等式的变化, 前面的集合 $\mathcal{T}(h) \subset T_K^2(G(x_0), DG(x_0)h)$ 被一可能大些的集合 $\mathcal{A}(h)$ 代替. 二阶近似集 $\mathcal{A}(h)$ 取得越小, 条件 (3.155) 与 (3.160) 就变得越强. 尤其, 若 $\mathcal{T}(h)$ 是一上二阶近似集合, 取 $\mathcal{A}(h) = \mathcal{T}(h)$, 条件 (3.155) 与 (3.160) 就变得最强了, 此种情形 (3.105) 与 (3.160) 的间隙就变为弱不等式与严格不等式间的差别, 正如非线性规划问题的情况. 这导致下述定义.

定义 3.85 称集合 K 在点 $y \in K$ 沿方向 $d \in T_K(y)$ 关于线性映射 $M : X \to Y$ 是外二阶正则的 (outer second order regular), 若对任何具有下述形式的 $y_n \in K$: $y_n := y + t_n d + \frac{1}{2} t_n^2 r_n$, 其中 $t_n \downarrow 0$, $r_n = M w_n + a_n$, $\{a_n\}$ 是 Y 中的收敛序列, $\{w_n\}$ 是 X 中的满足 $t_n w_n \to 0$ 的序列, 下述条件成立

$$\lim_{n \to \infty} \text{dist}(r_n, T_K^2(y, d)) = 0. \tag{3.161}$$

若 K 在 $y \in K$ 处沿每一方向 $d \in T_K(y)$ 相对于任何 M 均是外二阶正则的, 即 (3.161) 对任何满足 $t_n r_n \to 0$ 的序列 $y + t_n d + \frac{1}{2} t_n^2 r_n \in K$ 及 $d \in T_K(y)$ 成立, 称 K

在 y 处是二阶正则的 (second order regular), 若 K 在点 y 处对所有的 d 均是外上二阶正则的, 外二阶切集 $T_K^2(y,d)$ 与内二阶切集$T_K^{i,2}(y,d)$ 是相同的.

用于表示问题 (P) 的约束集合的集合 K 被设为是凸的. 然而, 注意到上述定义, 对于非凸集 K 上述定义也是有意义的. 在后续的分析中, 也考虑非凸的二阶正则的集合.

外二阶正则性的上述定义意味着外二阶切集 $T_K^2(y,d)$ 给出 K 在点 y 处沿方向 d 的一外二阶近似 (见定义 3.82). 类似地, 二阶正则性意味着内二阶切集$T_K^{i,2}(y,d)$ 给出了 K 在点 y 处沿方向 d 的外二阶近似. 因为外二阶切集包含在任何外二阶近似集中, 后者自动说明外二阶切集与内二阶切集重合. 换言之, (外) 二阶正则性意味着若 $y+td+\varepsilon(t)$ 在 K 中是与 d 相切的曲线, 即 $\varepsilon(t)=o(t)$, 则 $r(t):=\left(\frac{1}{2}t^2\right)^{-1}\varepsilon(t)$ 当 $t\downarrow 0$ 趋于 $T_K^{i,2}(y,d)$ (趋于 $T_K^2(y,d)$). 注意到 $r(t)$ 与其序列形式 $r_n:=r(t_n)$ 可以是无界的.

我们强调, 二阶正则集合对二阶分析 (second order analysis) 是适宜的概念. 对这样的集合, 相应的二阶必要条件与充分条件间没有间隙. 进一步, 我们后面将会看到这样的集合很自然地适合于最优值函数的二阶分析与最优解的稳定性分析的目的.

考虑下述二阶条件: 对任何 $h\in C(x_0)\setminus\{0\}$, 存在 $(\alpha,\lambda)\in\Lambda^g(x_0)$, 满足

$$D_{xx}^2L^g(x_0,\alpha,\lambda)(h,h)-\sigma(\lambda,\mathcal{T}(h))>0, \tag{3.162}$$

其中 $\mathcal{T}(h):=T_K^2(G(x_0),DG(x_0)h)$. 下述结果可由定理 3.45 与 3.83 得到.

定理 3.86 设 X 是有限维的. x_0 是问题 (P) 的可行点, 该点上的广义 Lagrange 乘子集合是非空的. 设对每一 $h\in C(x_0)$, 集合 K 在 $G(x_0)$ 处沿方向 $DG(x_0)h$ 相对于映射 $M:=DG(x_0)$ 是外二阶正则的. 则二阶条件 (3.162) 可推出在点 x_0 处的二阶增长条件. 若还有 x_0 处 Robinson 约束规范成立, 且对任何 $h\in C(x_0)$, 外二阶切集 $\mathcal{T}(h):=T_K^2(G(x_0),DG(x_0)h)$ 是凸的, 则二阶条件

$$\sup_{\lambda\in\Lambda(x_0)}\{D_{xx}^2L(x_0,\lambda)(h,h)-\sigma(\lambda,\mathcal{T}(h))\}>0,\quad \forall h\in C(x_0)\setminus\{0\} \tag{3.163}$$

是在点 x_0 处二阶增长条件成立的充分必要条件.

我们观察到, 若 K 在点 y 处沿方向 d 是外二阶正则的 (二阶正则的), 则必有外二阶切集$T_K^2(y,d)$ (内二阶切集 $T_K^{i,2}(y,d)$) 是非空的. 所以, 若 $T_K^2(y,d)$ 是空集, 则 K 在 y 处沿方向 d 不会是外二阶正则的. 例如, 考虑由例 3.29 构造的集合 S. 这个例子中, $T_S^2(x_0,h)=T_S^{i,2}(x_0,h)=\varnothing$, 因此 S 在 x_0 处不会是 (外) 二阶正则的. 在下述例子中, 构造凸集 S, 外与内二阶切集非空且相等, 但 S 不是二阶正则的.

例 3.87 在空间 $\mathbb{R}^3$ 中考虑集合

$$A := \{(x_1, x_2, x_3) : x_2 \geqslant x_1^2, x_1 \geqslant 0, x_3 = 0\},$$
$$B := \{(x_1, x_2, x_3) : x_3 \geqslant x_1^{3/2}, x_1 \geqslant 0, x_2 = 0\}.$$

令 $S := \mathrm{cl}[\mathrm{conv}(A \cup B)]$, $x_0 := (0,0,0), h := (1,0,0)$. 先来证明 $T_S^{i,2}(x_0, h) = T_S^2(x_0, h)$, 这时二阶切集与集合 $\mathcal{T} := \{(w_1, w_2, w_3) : w_2 \geqslant 2, w_3 \geqslant 0\}$ 相同.

集合 $\{(x_1, x_2, x_3) : x_2 \geqslant x_1^2, x_1 \geqslant 0, x_3 \geqslant 0\}$ 被包含在集合 S 中. 因为这一集合在 x_0 处沿方向 h 的二阶切集即同于 $\mathcal{T}$, 有 $\mathcal{T} \subset T_S^{i,2}(x_0, h)$. 令 (w_1, w_2, w_3) 是外二阶切集 $T_S^2(x_0, h)$ 中的一点. 这意味着对某一序列 $t_n \downarrow 0$, 存在 $\alpha_n, \beta_n, \bar{\alpha}_n, \bar{\beta}_n \in \mathbb{R}$ 及 $\tau_n \in [0,1]$, 满足 $\bar{\alpha}_n \geqslant \alpha_n, \bar{\beta}_n \geqslant \beta_n$, 量

$$\left\| \left(t_n + \frac{1}{2} t_n^2 w_1, \frac{1}{2} t_n^2 w_2, \frac{1}{2} t_n^2 w_3 \right) - (\tau_n \alpha_n + (1 - \tau_n) \beta_n, \tau_n \bar{\alpha}_n^2, (1 - \tau_n) \bar{\beta}_n^{3/2}) \right\|$$

的阶是 $o(t_n^2)$. 把上述条件按坐标展开有

$$\tau_n^{1/2} \left[\frac{1}{2} t_n^2 w_2 + o(t_n^2) \right]^{1/2} + (1 - \tau_n)^{1/3} \left[\frac{1}{2} t_n^2 w_3 + o(t_n^2) \right]^{2/3} \geqslant t_n + t_n^2 w_1 + o(t_n^2),$$

这导致

$$\tau_n^{1/2} \left(\frac{1}{2} w_2 \right)^{1/2} t_n \geqslant t_n + o(t_n).$$

因为 $\tau_n \leqslant 1$, 取极限可得 $w_2 \geqslant 2$. 直接可以验证必有 $w_3 \geqslant 0$. 结果有 $T_S^2(x_0, h) \subset \mathcal{T}$, 连同关系式 $\mathcal{T} \subset T_S^{i,2}(x_0, h) \subset T_S^2(x_0, h)$, 得到 $T_S^{i,2}(x_0, h) = T_S^2(x_0, h) = \mathcal{T}$.

现在考虑 $w_t := (0, 0, 2t^{-1/2}), t > 0$. 有 $x_0 + th + \frac{1}{2} t^2 w_t \in S$ 且 $t w_t \to 0$ 当 $t \downarrow 0$ 时成立. 另一方面, 对任何 $t > 0$ 均有 $\mathrm{dist}(w_t, T_S^2(x_0, h)) = 2$. 这表明 S 在 x_0 处沿方向 h 不是二阶正则的.

检验特定集合的二阶正则性可能是不容易的. 这一节剩余的部分对这一问题给出基本的分析, 后面还会回到这一问题上来, 比如证明正半定矩阵锥是二阶正则的. 由 3.34 命题, 可给出如下注记：若对每一 $d \in T_K(y)$, 有 $0 \in T_K^{i,2}(y, d)$, 则

$$T_K^{i,2}(y, d) = T_K^2(y, d) = T_{T_K(y)}(d),$$

因此, 此种情形下, K 在 y 处是二阶正则的. 比如, K 是多面集就是这种情况. 所以, 多面体集合总是二阶正则的.

下述命题表明, 二阶正则性在满足 Robinson 约束规范的二次连续可微变换之逆映射的作用下是被保持的.

命题 3.88 设 K 是 Y 中的闭凸子集, $G : X \to Y$ 是二次连续可微的映射. 若 Robinson 约束规范 (2.178) 于点 $x_0 \in G^{-1}(K)$ 处成立, K 在 $G(x_0)$ 处沿方向

$DG(x_0)h$ 相对于线性映射 $M := DG(x_0)$ 是外二阶正则的 (二阶正则的), 则 $G^{-1}(K)$ 在 x_0 处沿方向 h 是外二阶正则的 (二阶正则的).

证明　令 $x_n := x_0 + t_n h + \frac{1}{2}t_n^2 w_n \in G^{-1}(K)$ 满足 $t_n \downarrow 0, t_n w_n \to 0$. 由命题 3.33 和稳定性定理 (定理 2.87), 存在某一常数 $C > 0$, 对充分大的 n,

$$\begin{aligned}\operatorname{dist}(w_n, T^2_{G^{-1}(K)}(x_0,h)) &= \operatorname{dist}(w_n, DG(x_0)^{-1}[T^2_K(G(x_0), DG(x_0)h) - D^2G(x_0)(h,h)]) \\ &\leqslant C\operatorname{dist}(DG(x_0)w_n + D^2G(x_0)(h,h), T^2_K(G(x_0), DG(x_0)h)).\end{aligned}$$

$G(x_n)$ 的二阶展式为

$$G(x_n) = G(x_0) + t_n DG(x_0)h + \frac{1}{2}t_n^2(DG(x_0)w_n + D^2G(x_0)(h,h)) + o(t_n^2).$$

因为 $G(x_n) \in K$, 由 K 的外二阶正则性可得

$$\operatorname{dist}(DG(x_0)w_n + D^2G(x_0)(h,h), T^2_K(G(x_0), DG(x_0)h)) \to 0,$$

因此 $\operatorname{dist}(w_n, T^2_{G^{-1}(K)}(x_0,h)) \to 0$. $G^{-1}(K)$ 在 x_0 处沿方向 h 是外二阶正则的.

由命题 3.33 与 (3.59), (3.60) 可得, 如果

$$T^2_K(G(x_0), DG(x_0)h) = T^{i,2}_K(G(x_0), DG(x_0)h),$$

则外二阶切集 $T^2_{G^{-1}(K)}(x_0,h)$ 也等于内二阶切集$T^{i,2}_{G^{-1}(K)}(x_0,h)$. 结果, $G^{-1}(K)$ 的二阶正则性由 K 的二阶正则性可以得到. □

例如, 考虑集合

$$S := \{x \in X : g_i(x) \leqslant 0, i = 1,\cdots,p; h_j(x) = 0, j = 1,\cdots,q\},$$

它由有限多个约束定义. 设函数 g_i 与 h_j 是二阶连续可微的. 作为命题 3.88 的一个直接结果及多面集是二阶正则的这样的事实, 我们得到, 在满足 Mangasarian-Fromovitz 约束规范的每一 $x_0 \in S$ 处, 集合 S 均是二阶正则的.

设 $Y_1,\cdots,Y_n$ 是 Banach 空间, $K_i \subset Y_i, i = 1,\cdots,n$ 是闭子集. 考虑空间 $Y := Y_1 \times \cdots \times Y_n$ 与集合 $K := K_1 \times \cdots \times K_n \subset Y$. 不难看到, 由 $\mathcal{A}_{K_i}(y_i, d_i)$ 是 K_i 在 y_i 处沿方向 $d_i, i = 1,\cdots,n$ 的上二阶近似集, 可得

$$\mathcal{A}_K(y,d) := \mathcal{A}_{K_1}(y_1,d_1) \times \cdots \times \mathcal{A}_{K_n}(y_n,d_n) \tag{3.164}$$

是 K 在 $y = (y_1,\cdots,y_n)$ 处沿方向 $d := (d_1,\cdots,d_n)$ 的上二阶近似集合.

命题 3.89 设 $K_i \subset Y_i, i=1,\cdots,n$ 是闭集合. 设 K_i 在 $y_i \in K_i$ 沿方向 $d_i, i=1,\cdots,n$ 是二阶正则的. 则集合 $K := K_1 \times \cdots \times K_n$ 在 $y=(y_1,\cdots,y_n)$ 处沿方向 $d := (d_1,\cdots,d_n)$ 是二阶正则的.

证明 由于 $K_i \subset Y_i$ 是二阶正则的, $i=1,\cdots,n$, 可把 $T^2_{K_i}(y_i,d_i)$ 取作 K_i 在 y_i 处沿方向 d_i 的上二阶近似集. 由 (3.164) 与 (3.67) 得 $T^2_K(y,d)$ 是 K 在 y 处的沿方向 d 的一个上二阶近似集合, 这完成了证明. □

用相同的推证 (用 (3.69) 而不是 (3.67)) 可以证明, 若 K_1 在 $y_1 \in K_1$ 处沿方向 d_1 是外二阶正则的, K_2 在 $y_2 \in K_2$ 沿方向 d_2 是外二阶正则的, 则 $K_1 \times K_2$ 在 (y_1,y_2) 沿方向 (d_1,d_2) 是外二阶正则的.

命题 3.88 与 3.89 中的一个结果是下述的结论.

命题 3.90 设 $K_1,\cdots,K_n \subset Y$ 是闭凸集, 它们在点 $y_0 \in K_1 \cap \cdots \cap K_n$ 沿方向 $d \in T_{K_1}(y_0) \cap \cdots \cap T_{K_n}(y_0)$ 是二阶正则的. 存在 K_n 中的一点属于其余的 K_i 的内部, $i=1,\cdots,n-1$, 则 $K_1 \cap \cdots \cap K_n$ 在 y_0 处沿方向 d 是二阶正则的.

证明 应用命题 3.88, 其中 $G: Y \to Y \times \cdots \times Y$ 定义为 $G(y) := (y,\cdots,y)$, $K = K_1 \times \cdots \times K_n$. 不难看到 $G^{-1}(K) = K_1 \cap \cdots \cap K_n$. 对映射 G, Robinson 的约束规范可以像命题 3.36 那样验证. 再注意, 由命题 3.89 知 K 在点 $(y_0,\cdots,y_0)$ 沿方向 $(d,\cdots,d)$ 是二阶正则的, 这就证得结论. □

类似的推证可用于验证外二阶正则集合与二阶正则集之交集的外二阶正则性. 上述结果的一个简单的结论是, 若闭凸集 K 在 $y \in K$ 处是 (外) 二阶正则的, L 是 Y 中的一个线性闭子空间满足 $\operatorname{int}(K) \cap (y+L) \neq \varnothing$, 则 $K \cap (y+L)$ 在 y 处亦是 (外) 二阶正则的.

命题 3.91 令 $K_1,\cdots,K_n \subset Y$ 是闭集合, 且在 y_0 处沿方向 d 是外二阶正则的 (二阶正则的). 则集合 $K := \bigcup_{i=1}^{n} K_i$ 在 y_0 沿方向 d 是外二阶正则的 (二阶正则的).

证明 由命题 3.37 得

$$T^2_K(y_0,d) = \bigcup_{i=1}^{n} T^2_{K_i}(y_0,d). \tag{3.165}$$

用反证法.假设 K 在 y_0 处沿方向 d 不是外二阶正则的. 则存在序列 $y_n \in K$, $t_n \downarrow 0$ 满足 $y_n = y_0 + t_n d + \frac{1}{2}{t_n}^2 r_n$, $t_n r_n \to 0$, 且 $\operatorname{dist}(r_n, T^2_K(y_0,d)) \geqslant \varepsilon$, 其中 $\varepsilon > 0$ 是正数. 如有必要, 可取一子序列, 设所有的 y_n 属于集合 $K_1,\cdots,K_n$ 中的一个, 如 K_1. 因为 K_1 在 y_0 处沿方向 d 是外二阶正则的, 有 $\operatorname{dist}(r_n, T^2_{K_1}(y_0,d)) \to 0$. 由 (3.165) 得到矛盾. 这证得 K 的外二阶正则性质.

若 $T^2_{K_i}(y_0,d) = T^{i,2}_{K_i}(y_0,d), i=1,\cdots,n$, 则由 (3.165) 与 (3.76) 有

$$T^2_K(y_0,d) = \bigcup_{i=1}^{n} T^{i,2}_{K_i}(y_0,d) \subset T^{i,2}_K(y_0,d),$$

由于相反的包含关系总是成立的, 于是得到 $T_K^2(y_0,d)=T_K^{i,2}(y_0,d)$. 这就得到 K 的二阶正则性. □

命题 3.92 令 $K:=\{y:g(y)\leqslant 0\}$, 其中 $g:Y\to\overline{\mathbb{R}}$ 是下半连续的凸函数. 设存在 $\bar{y}$ 满足 $g(\bar{y})<0$ (Slater 条件), 令 $y_0,d\in Y$ 满足 $g(y_0)=0$ 且 $g^{\downarrow}(y_0,d)=0$. 若对任何具有形式 $y(t)=y_0+td+\dfrac{1}{2}t^2r(t),t\geqslant 0$ 的路径$y(t)\in K$, 其中 $r(t)$ 满足当 $t\downarrow 0$ 时 $tr(t)\to 0$, 不等式

$$\limsup_{t\downarrow 0}\ g_-^{\downarrow\downarrow}(y_0;d,r(t))\leqslant 0 \tag{3.166}$$

成立, 则 K 在 y_0 处沿方向 d 是外二阶正则的. 若还有 $g(\cdot)$ 是连续的, 则相反的结论也是成立的, 即由 K 在 y_0 处沿方向 d 的外二阶正则性可推出 (3.166).

证明 由命题 2.61 可得

$$T_K(y_0)=\{d:g^{\downarrow}(y_0,d)\leqslant 0\},$$

从而由 $g^{\downarrow}(y_0,d)=0$ 可得 $d\in T_K(y_0)$.

设 (3.166) 成立. 考虑序列 $y_n:=y_0+t_nd+\dfrac{1}{2}t_n^2r_n\in K$ 满足 $t_n\downarrow 0$, 且 $t_nr_n\to 0$. 选取 $\alpha>0$, 固定 $n\in\mathbb{N}$, 令 $w_{\alpha,n}:=r_n+\alpha(\bar{y}-y_0)$. 则由 g 的凸性, 对 $t\geqslant 0$ 充分小, 有

$$g\Big(y_0+td+\frac{1}{2}t^2w_{\alpha,n}\Big)\leqslant\Big(1-\frac{1}{2}\alpha t^2\Big)\gamma(t,r_n)+\frac{1}{2}\alpha t^2g(\bar{y}), \tag{3.167}$$

其中

$$\gamma(t,r):=g\Big(y_0+t\Big(1-\frac{1}{2}\alpha t^2\Big)^{-1}d+\frac{1}{2}t^2\Big(1-\frac{1}{2}\alpha t^2\Big)^{-1}r\Big).$$

同命题 3.30 的证明一样 (确切说, 由 (3.55) 及其随后的推导), 得到

$$g_-^{\downarrow\downarrow}(y_0;d,w_{\alpha,n})\leqslant g_-^{\downarrow\downarrow}(y_0;d,r_n)+\alpha g(\bar{x}). \tag{3.168}$$

因为 $g(\bar{y})<0$, 由 (3.166) 与 (3.168) 得, 存在 n_0(依赖于 α) 满足对所有的 $n\geqslant n_0$, $g_-^{\downarrow\downarrow}(y_0;d,w_{\alpha,n})<0$, 因此由命题 3.30 有 $w_{\alpha,n}\in T_K^2(y_0,d)$. 结果

$$\limsup_{n\to\infty}\operatorname{dist}(r_n,T_K^2(y_0,d))\leqslant\ \alpha\|\bar{y}-y_0\|.$$

由于 α 可以取得任意小, 得到

$$\lim_{n\to\infty}\ \operatorname{dist}(r_n,T_K^2(y_0,d))=0,$$

因此 K 在 y_0 处沿方向 d 是外二阶正则的.

相反地, 设 K 在 y_0 处沿方向 d 是外二阶正则的, 满足 $g^{\downarrow}(y_0,d)=0$ 且 $g(\cdot)$ 是连续的. 由于 g 是连续的, $g^{\downarrow}(y_0,\cdot)=g'(y_0,\cdot)$, $g_{-}^{\downarrow\downarrow}(y_0,d,\cdot)=g_{-}''(y_0,d,\cdot)$. 令 $t_n\downarrow 0$ 是 (3.166) 的上极限达到极限的序列, $r_n:=r(t_n)$. 置 $\varepsilon_n:=\mathrm{dist}\ (r_n,T_K^2(y_0,d))+n^{-1}$. 由外二阶正则性可得 $\varepsilon_n\to 0$. 选取 $\tilde{r}_n\in T_K^2(y_0,d)$ 满足 $\|r_n-\tilde{r}_n\|<\varepsilon_n-\dfrac{1}{2}n^{-1}$. 则对给定的 n, 存在序列 $\tau_k\downarrow 0$, 满足

$$y_0+\tau_k d+\frac{1}{2}\tau_k^2\tilde{r}_n+o(\tau_k^2)\in K,$$

则对 k 充分大, 有

$$y_0+\tau_k d+\frac{1}{2}\tau_k^2 r_n\in K+\frac{1}{2}\varepsilon_n\tau_k^2 B_Y.$$

结果, 对 $\alpha>0, w_{\alpha,n}:=r_n+\alpha(\bar{y}-y_0)$, 有

$$\begin{aligned}
y_0+\tau_k d+\frac{1}{2}\tau_k^2 w_{\alpha,n}&=\left(1-\frac{1}{2}\alpha\tau_k^2\right)\left(y_0+\tau_k d+\frac{1}{2}\tau_k^2 r_n\right)+\frac{1}{2}\alpha\tau_k^2\left(\bar{y}+\tau_k d+\frac{1}{2}\tau_k^2 r_n\right)\\
&\subset\left(1-\frac{1}{2}\alpha\tau_k^2\right)\left(K+\frac{1}{2}\varepsilon_n\tau_k^2 B_Y\right)+\frac{1}{2}\alpha\tau_k^2\left(\bar{y}+\tau_k d+\frac{1}{2}\tau_k^2 r_n\right)\\
&=\left(1-\frac{1}{2}\alpha\tau_k^2\right)K+\frac{1}{2}\alpha\tau_k^2\Big[\bar{y}+\tau_k d+\frac{1}{2}\tau_k^2 r_n\\
&\quad+\left(1-\frac{1}{2}\alpha\tau_k^2\right)\varepsilon_n\alpha^{-1}B_Y\Big].
\end{aligned}$$

由于 $g(\bar{y})<0$, g 是连续的, 有 $\bar{y}\in\mathrm{int}(K)$. 因此, 对 n 充分大, $\bar{y}+2\varepsilon_n\alpha^{-1}B_Y\subset K$. 由 K 的凸性及上述的推导, 对所有的 k 与充分大的 n, $y_0+\tau_k d+\dfrac{1}{2}\tau_k^2 w_{\alpha,n}\in K$, 因此有 $w_{\alpha,n}\in T_K^2(y_0,d)$. 由命题 3.30 得 $g_{-}^{\downarrow\downarrow}(y_0;d,w_{\alpha,n})\leqslant 0$. 由于 g 在 y_0 处是连续的, 它是局部 Lipschitz 连续的. 因此, $g_{-}''(y_0;d,\cdot)$ 是全局 Lipschitz 连续的, Lipschitz 常数与 g 在 y_0 处的相同, 设为 L, 则

$$g_{-}''(y_0;d,r_n)\leqslant L\|w_{\alpha,n}-r_n\|=\alpha L\ \|\bar{y}-y_0\|,$$

因此得到

$$\limsup_{t\downarrow 0}\ g_{-}''(y_0;d,r(t))=\lim_{n\to\infty}\ g_{-}''(y_0;d,r_n)\leqslant\alpha L\ \|\bar{y}-y_0\|.$$

因为 α 可取得任意小, (3.166) 得证. □

若 $g(\cdot)$ 是连续的, $g(y_0)=0, g'(y_0,d)<0$, 则 $T_K^2(y_0,d)=Y$. 所以, 此种情况, K 在 y_0 沿方向 d 显然是二阶正则的.

现在导出验证条件 (3.166) 的一些准则. 观察到, 当不等式

$$g\left(y_0+td+\frac{1}{2}t^2 r(t)\right)\geqslant g(y_0)+tg^{\downarrow}(y_0,d)+\frac{1}{2}t^2 g_{-}^{\downarrow\downarrow}(y_0;d,r(t))+o(t^2)\tag{3.169}$$

对 $t \geqslant 0$ 成立时, 其中 $r(t)$ 满足当 $t \downarrow 0$ 时 $tr(t) \to 0$, 则这一条件 (即 (3.166)) 成立. 例如, 如果 g 是二次连续可微的, 则这一条件成立. 在下一节将讨论满足上述条件的函数 $g(y)$, 这些函数被称为二阶上图正则的 (second order epiregular).

令 $g_i : Y \to \overline{\mathbb{R}}, i = 1, \cdots, n$ 是下半连续函数, $K_i := \{y := g_i(y) \leqslant 0\}$. 有

$$K := \bigcap_{i=1}^{n} K_i = \{y : \max_{1 \leqslant i \leqslant n} g_i(y) \leqslant 0\} = \{y \in Y : g_i(y) \leqslant 0, i = 1, \cdots, n\}.$$

所以, 由命题 3.90, 若还有 g_i 是凸的, 集合 K_i 是二阶正则的, 且在 K_n 存在一点属于其余的 K_i 的内部, 则上述集合 K 也是二阶正则的.

类似地, 由命题 3.91, 若 K_i 是二阶正则的, 则集合 $\bigcup_{i=1}^{n} K_i = \{y : \min_{1 \leqslant i \leqslant n} g_i(y) \leqslant 0\}$ 也是二阶正则的.

3.3.4 函数的二阶正则性

将 3.3.3 节发展起来的对集合的二阶正则性的概念推广到增广的 (不必是凸的) 实值函数上是非常重要的. 很自然, 将对集合的二阶正则性的概念用于 g 的上图 $K := \text{epi } g$, 来解决函数的二阶正则性这一问题.

定义 3.93 设 $g : Y \to \overline{\mathbb{R}}$ 是增广实值函数, 它在 y_0 处取有限值. 称 $g(\cdot)$ 在点 y_0 沿方向 d 是 (外) 二阶正则的, 若 $g_-^{\downarrow}(y_0, d)$ 是有限的, 且集合 $K := \text{epi } g$ 在点 $(y_0, g(y_0))$ 处沿方向 $(d, g_-^{\downarrow}(y_0, d))$ 是 (外) 二阶正则的. 称 $g(\cdot)$ 在 y_0 处是 (外) 二阶正则的, 若集合 $K := \text{epi } g$ 在点 $(y_0, g(y_0))$ 处是 (外) 二阶正则的.

应该注意到, 将二阶正则性的概念翻译成函数的语言时应该格外小心. 若对一特定的方向, 下方向上图导数$g_-^{\downarrow}(y_0, d)$ 是无穷的, 则谈方向 $(d, g_-^{\downarrow}(y_0, d))$ 就没有意义了. 另一方面, 当视 $K := \text{epi } g$ 为一集合时, 可以考虑竖直的方向. 若函数 g 是凸的, 在 y_0 处是连续的, 则 $g_-^{\downarrow}(y_0, d) = g'(y_0, d)$ 且对所有的 d, 这些方向导数是有界的. 此种情形, 可定义 g 沿每一方向 $d \in Y$ 的 (外) 二阶正则性. 此种情况, 若 (d, γ) 满足 $\gamma > g'(y_0, d)$, 则相应的二阶切集 $T_K^2((y_0, g_0), (d, \gamma))$ 与空间 $Y \times \mathbb{R}$ 重合, 这意味着集合 K 沿方向 (d, γ) 是二阶正则的. 所以, 在凸的连续的情形, 只需要验证 K 沿形式为 $(d, g'(y_0, d))$ 的方向的二阶正则性.

在应用的时候, 直接处理函数 g 而不是其上图是适宜的. 从这一角度看, 下述概念是有用途的.

定义 3.94 设 $g : Y \to \overline{\mathbb{R}}$ 是增广实值函数, 在 y_0 处取有限值. 称 $g(\cdot)$ 在点 y_0 处沿 d 方向是二阶上图正则的 (second order epiregular), 若 $g_-^{\downarrow}(y_0, d)$ 是有限的且下述条件成立: 对任何满足 $tr(t) \to 0, t \downarrow 0$ 的每一路径 $r : \mathbb{R}_+ \to Y$, 下述不等式对 $t \geqslant 0$ 成立

$$g\Big(y_0 + td + \frac{1}{2}t^2 r(t)\Big) \geqslant g(y_0) + tg_-^{\downarrow}(y_0, d) + \frac{1}{2}t^2 g_-^{\downarrow\downarrow}(y_0; d, r(t)) + o(t^2). \qquad (3.170)$$

很清楚, 若 g 是二次连续可微的, 则它是二阶上图正则的. 后面将会看到, 二阶上图正则性对各种情况下的最优值函数是可以验证的.

命题 3.95 设 $g(y)$ 是增广实值函数, 在 y_0 处取有限值. 若 $g_-^\downarrow(y_0,d)$ 是有限的, g 在 y_0 处沿方向 d 是二阶上图正则的, 则 g 在 y_0 处沿方向 d 是外二阶正则的. 进一步, 若 g 在 y_0 处是方向可微的且在 y_0 的邻域内是 Lipschitz 连续的, 则 g 在 y_0 处沿方向 d 是外二阶正则的充要条件是 g 在 y_0 处沿方向 d 为二阶上图正则的.

证明 设 g 在 y_0 处沿方向 d 是二阶上图正则的. 令 $t_n \downarrow 0$, $(r_n, c_n) \in Y \times \mathbb{R}$ 满足 $t_n(r_n, c_n) \to 0$, 且

$$(y_0, g(y_0)) + t_n(d, g_-^\downarrow(y_0,d)) + \frac{1}{2}t_n^2(r_n, c_n) \in \mathrm{epi}g, \tag{3.171}$$

即

$$g(y_0) + t_n g_-^\downarrow(y_0,d) + \frac{1}{2}t_n^2 c_n \geqslant g\Big(y_0 + t_n d + \frac{1}{2}t_n^2 r_n\Big). \tag{3.172}$$

由 (3.170) 可推出 $c_n' \geqslant g_-''(y_0;d,r_n)$, 其中 $\{c_n'\}$ 是满足 $c_n - c_n' \to 0$ 的某一序列. 则由公式 (3.86), 即 epi g 的外二阶切集的公式, 有 (r_n, c_n') 属于 K 的相应的外二阶切集, 这就得到 K 的外二阶正则性质.

相反地, 设 g 在 y_0 点是外二阶正则的, 方向可微并 Lipschitz 连续的. 令 $t_n \downarrow 0, r_n \in Y$ 满足 $t_n r_n \to 0$. 定义

$$c_n := \frac{g\Big(y_0 + t_n d + \frac{1}{2}t_n^2 r_n\Big) - g(y_0) - t_n g'(y_0,d)}{\frac{1}{2}t_n^2}. \tag{3.173}$$

注意, 对这样定义的 c_n, (3.172) 满足且成等式. 由于 g 在 y_0 是方向可微的且在 y_0 处是 Lipschitz 连续的, 有 g 在 y_0 是 Hadamard 方向可微的. 这得到 $t_n c_n \to 0$. 由于 g 是外二阶正则的, 由公式 (3.86) 得, 存在序列 $\{(r_n', c_n')\}$ 满足 $(r_n, c_n) - (r_n', c_n') \to 0$ 且 $c_n' \geqslant g_-^{\downarrow\downarrow}(y_0;d,r_n')$. 这得到

$$g\Big(y_0 + t_n d + \frac{1}{2}t_n^2 r_n\Big) \geqslant g(y_0) + t_n g_-^\downarrow(y_0,d) + \frac{1}{2}t_n^2 g_-^{\downarrow\downarrow}(y_0;d,r_n') + o(t_n^2). \tag{3.174}$$

由于 g 在 y_0 附近是 Lipschitz 连续的, $g_-^{\downarrow\downarrow}(y_0;d,\cdot)$ 是 Lipschitz 连续的, 有不等式 (3.174) 右端的 r_n' 可替换为 r_n. 这就证得 g 的二阶上图正则性. □

上述命题表明, 二阶上图正则性是上二阶正则性的充分条件, 对凸的连续函数而言, 外二阶正则性与二阶上图正则性是一样的. 为理解凸的非连续函数的外二阶正则性与二阶上图正则性的区别, 考虑一个指示函数$g(\cdot) = I_S(\cdot)$ 的例子, 其中 S 是 Y 的非空闭凸子集, $y \in S$. 由于 I_S 的上图由 $S \times \mathbb{R}_+$ 给出, 有 I_S 在 y 处外二阶正则的充要条件是集合 S 在 y 是外二阶正则的 (见命题 3.89 及这一命题后的讨论).

考虑方向 $d \in T_S(y)$. 则 $g_-^{\downarrow}(y,d)=0$, 由 (3.86) 得 $g_-^{\downarrow\downarrow}(y\,;d,\cdot)=I_{T_S^2(y,d)}(\cdot)$. 因此, 不等式 (3.170) 成立当且仅当 $y+td+\frac{1}{2}t^2r(t)\in S$ 可推出 $r(t)\in T_S^2(y,d)$. 然而, 即使 S 在 y 处是二阶正则的, 这个包含关系也可能不成立. 这就表明, 对非连续函数而言, 二阶上图正则性是外二阶正则性的充分条件, 但一般来说不是必要的条件.

下述结果是由命题 3.88 给出的关于集合相应结论的一个直接结果.

命题 3.96 设 $G:X\to Y$ 是二次连续可微映射, $f:Y\to\overline{\mathbb{R}}$ 是下半连续的凸函数, 在 $y_0:=G(x_0)$ 处取有限值, $h\in X$. 设 f 在 y_0 处沿方向 $DG(x_0)h$ 是外二阶正则的, Robinson 约束规范(3.87) 成立. 则复合函数 $g:=f\circ G$ 在 x_0 沿方向 h 是外二阶正则的.

不难验证, Lipschitz 连续的二阶上图正则 (不必是凸的) 函数与二次连续可微映射的复合是二阶上图正则的. 由命题 3.95, 对方向可微的 Lipschitz 连续函数, 二阶上图正则性的概念与外二阶正则性的概念是相同的.

命题 3.97 设 $G:X\to Y$ 是二阶连续可微映射, $f:Y\to\mathbb{R}$ 在 $y_0:=G(x_0)$ 附近为 Lipschitz 连续的函数, 它在 y_0 处是二阶方向可微的且是二阶上图正则的. 则复合函数 $g:=f\circ G$ 在 x_0 处是二阶上图正则的.

证明 由于 f 是 Lipschitz 连续的, 则 f 在 y_0 处是 Hadamard 二阶方向可微的. 由命题 2.53, 复合函数 g 在 x_0 处是二阶方向可微的, 相应的链式法则 (2.82) 成立. 则用 G 的二阶 Taylor 展开式, 直接可以验证相应的不等式 (3.170) 成立. □

例 3.98 设 $\lambda_1(A)\geqslant\cdots\geqslant\lambda_p(A)$ 是 $p\times p$ 对称矩阵 A 的特征值, 令 $i\in\{1,\cdots,p\}$, 称 $\lambda_i(A)$ 是 A 的 r 重数 (multiplicity) 的首项 (leading) 特征值, 若 $\lambda_{i-1}(A)>\lambda_i(A)=\cdots=\lambda_{i+r-1}(A)>\lambda_{i+r}(A)$. 由定义, 最大特征值 $\lambda_1(A)$ 是首项的特征值. 我们证明, 若 $\lambda_i(A)$ 是 A 的首项的特征值, 则函数 $\lambda_i(\cdot)$ 在 $A\in\mathcal{S}^p$ 处是二阶上图正则的.

首先证明, 存在映射 $\Xi:\mathcal{S}^p\to\mathcal{S}^r$ 满足: (i) Ξ 在 A 的邻域内是二次连续可微的; (ii) $\Xi(A)=\alpha I_r$, 其中 $\alpha:=\lambda_i(A)$, I_r 是 $r\times r$ 单位矩阵; (iii) $D\Xi(A):\mathcal{S}^p\to\mathcal{S}^r$ 是映上的; (iv) 对 A 的邻域中的所有 X, 特征值 $\lambda_1(\Xi(X))\geqslant\cdots\geqslant\lambda_r(\Xi(X))$ 与特征值 $\lambda_i(X)\geqslant\cdots\geqslant\lambda_{i+r-1}(X)$ 重合. 这样一个映射的下述构造类似于例 3.140 中的构造.

用 $L(X)$ 记对应特征值 $\lambda_i(X)\geqslant\cdots\geqslant\lambda_{i+r-1}(X)$ 的特征空间 (eigenspace), 令 $P(X)$ 是到 $L(X)$ 上的直交投影矩阵. 令 E_0 是 (固定的)$p\times r$ 矩阵, 它的列是直交的且张成空间 $L(A)$. 注意到, 在 A 的充分小邻域中, $P(X)$ 是 X 的二阶连续可微 (事实上, 甚至是解析) 的函数. 结果, $F(X):=P(X)E_0$ 在 A 的一邻域中亦是 X 的二阶连续可微函数, 进一步, $F(A)=E_0$. 对充分接近 A 的所有的 X, $F(X)$ 的秩是 r, 即它的列向量是线性无关的. 令 $U(X)$ 是矩阵, 其列由 $F(X)$ 的列经过 Gram-Schmidt 正交化过程得到. 矩阵 $U(X)$ 是有定义的, 且在 X 的附近是二阶连续可微的 (事实

上, 是解析的), 同时还满足下述条件: $U(A) = E_0$, $U(X)$ 的列空间与 $E(X)$ 的列空间重合, 且 $U(X)^{\mathrm{T}}U(X) = I_r$. 考虑映射 $\Xi(X) := U(X)^{\mathrm{T}}XU(X)$. 映射 Ξ 是二阶连续可微的 (事实上, 甚至是解析的), $D\Xi(A)$ 是映上的. 还有 $\Xi(A) := E_0^{\mathrm{T}}AE_0 = \alpha I_r$, 由于 $U(X)$ 是正交的且张成 X 的相应的特征空间, 性质 (iv) 也是成立的.

令 $\lambda_i(A)$ 是重数为 r 的 A 的首项特征值. 考虑满足 $t \downarrow 0$ 时 $tW_t \to 0$ 的形式为 $X(t) := A + tH + \frac{1}{2}t^2W_t$ 的路径 $X(t) \in \mathcal{S}^p$. 因为对 $t > 0$ 充分小, $\lambda_i(X(t)) = \lambda_1(\Xi(X(t)))$, 有

$$\Xi(X(t)) = \alpha I_r + tD\Xi(A)H + \frac{1}{2}t^2[D\Xi(A)W_t + D^2\Xi(A)(H,H)] + o(t^2),$$

且因函数 $\lambda_1(\cdot) : \mathcal{S}^r \to \mathbb{R}$ 是 Lipschitz 连续的, $\lambda_1(\alpha I_r + B) = \alpha + \lambda_1(B)$ 对任何 $B \in \mathcal{S}^r$ 成立, 得到

$$\lambda_i(X(t)) = \alpha + t\lambda_1[D\Xi(A)H + \frac{1}{2}t(D\Xi(A)W_t + D^2\Xi(A)(H,H))] + o(t^2). \tag{3.175}$$

由函数 $\lambda_1(\cdot)$ 是凸的, 有

$$\lambda_i(X(t)) \geqslant \alpha + t\lambda_1(D\Xi(A)H) + \frac{1}{2}t^2\lambda_1'(D\Xi(A)H, D\Xi(A)W_t + D^2\Xi(A)(H,H)) + o(t^2). \tag{3.176}$$

因 $\alpha := \lambda_i(A)$, 由 (3.175) 可得

$$\lambda_i'(A,H) = \lambda_1(D\Xi(A)H), \tag{3.177}$$

$$\lambda_i''(A;H,W) = \lambda_1'(D\Xi(A)H, D\Xi(A)W + D^2\Xi(A)(H,H)). \tag{3.178}$$

进一步, 由 (3.176) 可推出不等式

$$\lambda_i(X(t)) \geqslant \lambda_i(A) + t\lambda_i'(A,H) + \frac{1}{2}t^2\lambda_i''(A;H,W_t) + o(t^2) \tag{3.179}$$

对 $t \geqslant 0$ 成立, 因此得到 $\lambda_i(\cdot)$ 在 A 处的二阶上图正则性.

若 $G : \mathbb{R}^n \to \mathcal{S}^p$ 是二次连续可微映射, $\bar{x} \in \mathbb{R}^n$, $\lambda_i(A)$ 是 $A := G(\bar{x})$ 的首项特征值, 则由命题 3.97 可得复合函数 $\lambda_i(G(x))$ 在 $\bar{x}$ 处是二阶上图正则的.

3.3.5 二阶次导数

这一节我们从稍微不同的角度出发讨论二阶最优性条件. 令 X 是 Banach 空间, $f : X \to \overline{\mathbb{R}}$ 是正常的 (不必凸的) 函数. 这意味着对所有的 $x \in X$, 有 $f(x) > -\infty$, f 的定义域非空. 考虑优化问题

$$\text{(P)} \qquad \min_{x \in X} f(x). \tag{3.180}$$

显然, 上述的极小化问题可限制在 f 的定义域上极小化函数 f. 称 $x \in X$ 是问题 (P) 的可行点, 若 $x \in \mathrm{dom} f$.

用下方向上图导数 (见 2.2.3 节) 不难给出问题(P) 的一阶最优性条件.

命题 3.99　设 x_0 是 (P) 的可行点. 则 (i) 若 x_0 是 (P) 的局部最优解, 则对所有的 $h \in X, f_-^\downarrow(x_0, h) \geqslant 0$; (ii) 若还有空间 X 是有限维的, 则一阶增长条件在 x_0 处成立当且仅当

$$f_-^\downarrow(x_0, h) > 0, \quad \forall h \in X \backslash \{0\}. \tag{3.181}$$

证明　设 $x_0 \in \mathrm{dom} f$ 是 (P) 的局部最优解. 则对接近 x_0 的所有 x, $f(x) \geqslant f(x_0)$, 因此由 f 的下方向上图导数定义 (2.68) 得到 $f_-^\downarrow(x_0, h) \geqslant 0$. 若一阶增长条件在 x_0 处成立, 则存在 $c > 0$, 对充分小的 $t \geqslant 0$, 有 $f(x_0 + th) - f(x_0) \geqslant ct\|h\|$. 此种情况, 对任何 $h \neq 0$, 有 $f_-^\downarrow(x_0, h) \geqslant c\|h\| > 0$.

假设当 X 是有限维空间时, 条件 (3.181) 成立但一阶增长条件不成立. 则存在 $X \backslash \{x_0\}$ 中的收敛到的 x_0 的序列 $\{x_n\}$, 满足

$$f(x_n) \leqslant f(x_0) + o(\|x_n - x_0\|). \tag{3.182}$$

考虑 $h_n := (x_n - x_0)/\|x_n - x_0\|$. 由于 X 是有限维的, 若有必要可取子列, 设 h_n 收敛到向量 $h \neq 0$. 则由 (3.182) 得 $f_-^\downarrow(x_0, h) \leqslant 0$, 这与 (3.181) 矛盾. □

现在考虑如下的所谓的二阶次导数(second order subderivative)

$$d^2 f(x\ |\alpha)(h) := \liminf_{\substack{t \downarrow 0 \\ h' \to h}} \frac{f(x + th') - f(x) - t\langle \alpha, h' \rangle}{\frac{1}{2}t^2}. \tag{3.183}$$

注意到, 二阶次导数$d^2 f(x|\alpha)(\cdot)$ 在可行点 $x \in \mathrm{dom} f$ 上定义, 且依赖于 $\alpha \in X^*$. 注意到 $d^2 f(x|\alpha)(\cdot)$ 定义为相应的比率函数的下上图-极限, 因此是下半连续的.

对 $\alpha = 0$, 二阶次导数$d^2 f(x|0)(\cdot)$ 的形式很像下方向上图导数$f_-^\downarrow(x, \cdot)$, 当然, 其差别是在 (3.183) 中分母中的 $\frac{1}{2}t^2$ 被 $f_-^\downarrow(x, \cdot)$ 定义中的 t 代替. 下述结果可如命题 3.99 中的证明那样加以证明.

命题 3.100　设 x_0 是问题 (P) 的可行点. 则 (i) 若 x_0 是 (P) 的局部最优解, 则对所有的 $h \in X, d^2 f(x_0\ |0)(h) \geqslant 0$; (ii) 若还有空间 X 是有限维的, 则二阶增长条件在 x_0 处成立当且仅当

$$d^2 f(x_0\ |0)(h) > 0, \quad \forall h \in X \backslash \{0\}. \tag{3.184}$$

由定义不难看到, 若对某一 h, $f_-^\downarrow(x_0, h) > 0$, 则 $d^2 f(x_0\ |0)(h) = +\infty$. 因此, 假定一阶条件 $f_-^\downarrow(x_0, \cdot) \geqslant 0$ 时, 二阶条件 (3.184) 应该仅对 $h \in C(x_0) \backslash \{0\}$ 验证, 其中,

$$C(x_0) := \{h \in X : f_-^\downarrow(x_0, h) = 0\}. \tag{3.185}$$

将 $C(x_0)$ 称为问题 (P) 在 x_0 处的临界锥 (critical cone). 注意到, 如果 f 不是凸的, 则锥 $C(x_0)$ 不一定是凸的.

理解二阶次导数$d^2f(x\,|\alpha)(h)$ 与 (抛物的) 二阶下上图导数 $f_-^{\downarrow\downarrow}(x;h,w)$ 的关系是重要的.

命题 3.101 *设 x_0 是问题 (P) 的可行点, 对给定的 $h\in X$, 令 $\alpha\in X^*$ 满足 $\langle\alpha,h\rangle=f_-^{\downarrow}(x,h)$. 则*

$$\inf_{w\in X}\{f_-^{\downarrow\downarrow}(x;h,w)-\langle\alpha,w\rangle\}\geqslant d^2f(x\,|\alpha)(h). \tag{3.186}$$

证明 令 $w_n\to w, t_n\downarrow 0$, 在 $d^2f(x\,|\alpha)(h)$ 的定义 (3.183) 中, 取 $h_n':=h+\frac{1}{2}t_nw_n$. 得

$$d^2f(x\,|\alpha)(h)\leqslant\liminf_{n\to\infty}\frac{f\Big(x+t_nh+\frac{1}{2}t_nw_n^2\Big)-f(x)-t_nf_-^{\downarrow}(x,h)-\frac{1}{2}t_n^2\langle\alpha,w_n\rangle}{\frac{1}{2}t_n^2}.$$

于是对任何 $w\in X$, 有 $d^2f(x\,|\alpha)(h)\leqslant f_-^{\downarrow\downarrow}(x,h,w)-\langle\alpha,w\rangle$. 在上式 (不等式) 的右端取所有 $w\in X$ 的下确界, 即得 (3.186). □

定义 3.102 *函数 f 在可行点 x 沿方向 h, 对 α 是抛物正则的 (parabolically regular), 若 (3.186) 中的等式成立, 即*

$$\inf_{w\in X}\{f_-^{\downarrow\downarrow}(x;h,w)-\langle\alpha,w\rangle\}=d^2f(x\,|\alpha)(h). \tag{3.187}$$

注意, (3.187) 意味着 $d^2f(x\,|\alpha)(h)=-\phi^*(\alpha)$, 其中 $\phi^*(\cdot)$ 是函数 $\phi(\cdot):=f_-^{\downarrow\downarrow}(x\,;h,\cdot)$ 的共轭函数. 因此, 上述抛物正则性揭示了二阶次导数与二阶方向 (抛物的) 上图导数间的对偶关系.

命题 3.103 *设 x 是 (P) 的可行点, $h\in X,\alpha\in X^*$ 满足 $d^2f(x\,|\alpha)(h)>-\infty,\langle\alpha,h\rangle=f_-^{\downarrow}(x,h)$. 若函数 f 在 x 处沿方向 h 是外二阶正则的, 则 f 在 x 处沿方向 h, 对 α 是抛物正则的.*

证明 若 $d^2f(x\,|\alpha)(h)=+\infty$, 则由 (3.186) 得 (3.187) 是成立的. 因此可设 $d^2f(x\,|\alpha)(h)$ 是有限的. 注意, 由假设 $\langle\alpha,h\rangle=f_-^{\downarrow}(x,h)$, 有 $f_-^{\downarrow}(x,h)$ 是有限的.

设 f 在 x 处沿方向 h 是外二阶正则的. 令 $h_n\to h,t_n\downarrow 0$ 是使得 (3.183) 的右端极限取到的序列. 考虑 $w_n:=\left(\frac{1}{2}t_n\right)^{-1}(h_n-h)$, 即 $h_n=h+\frac{1}{2}t_nw_n, x_n:=x+t_nh+\frac{1}{2}t_n^2w_n$. 则 $t_nw_n\to 0$, 且

$$d^2f(x\,|\alpha)(h)=\lim_{n\to\infty}\frac{f(x_n)-f(x)-t_nf_-^{\downarrow}(x,h)-\frac{1}{2}t_n^2\langle\alpha,w_n\rangle}{\frac{1}{2}t_n^2}. \tag{3.188}$$

考虑

$$c_n := \frac{f(x_n) - f(x) - t_n f_-^{\downarrow}(x,h)}{\frac{1}{2}t_n^2}.$$

因为 $t_n w_n \to 0$, 注意到 (3.188) 及 $d^2 f(x\,|\alpha)(h)$ 是有限的, 有 $t_n c_n \to 0$. 由 f 的上图的外二阶切集的公式 (3.86) 及 f 的外二阶正则性, 存在序列 $\{w_n'\}$ 满足 $w_n - w_n' \to 0$ (与命题 3.95 的证明比较), 有

$$f(x_n) \geqslant f(x) + t_n f_-^{\downarrow}(x,h) + \frac{1}{2}t_n^2 f_-^{\downarrow\downarrow}(x;h,w_n') + o(t_n^2), \tag{3.189}$$

将 (3.189) 的右端代替 (3.188) 中的 $f(x_n)$, 得到

$$d^2 f(x\,|\alpha)(h) \geqslant \liminf_{n\to\infty}\{f_-^{\downarrow\downarrow}(x;h,w_n') - \langle\alpha, w_n'\rangle - \langle\alpha, w_n - w_n'\rangle\}^{①}$$

由于 $\langle\alpha, w_n - w_n'\rangle \to 0$, 这就得到 $d^2 f(x\,|\alpha)(h)$ 大于或等于 (3.187) 的左端, 因此等号成立. □

注意, 若 f 是凸的, $\alpha \in \partial f(x)$, 则 (3.183) 中比率中的分子是非负的, 从而 $d^2 f(x\,|\alpha)(h) \geqslant 0$. 所以, 若 f 是凸函数时, 则上述命题中的条件 $d^2 f(x\,|\alpha)(h) > -\infty$ 是多余的.

上述命题表明, 对满足 $\langle\alpha, h\rangle = f_-^{\downarrow}(x,h)$ 的 $\alpha \in X^*$, 外二阶正则性是抛物正则性的充分条件. 为了说明这一条件不是必要的, 考虑例子 $f(x) := x^{4/3}, x \in \mathbb{R}$, 其中 $x = 0, h = 1, \alpha = 0$. 此种情形, 等式 (3.187) 成立, 其公共值是 $+\infty$, 而 f 在 $x = 0$ 处不是外二阶正则的.

用命题 3.103 中的结果, 由相应的 (抛物) 二阶上图导数的链式法则, 可以导出二阶次导数的链式法则.

命题 3.104　设 X 与 Y 是 Banach 空间, $F: X \to Y$ 是二阶连续可微函数映射, $g: Y \to \overline{\mathbb{R}}$ 是下半连续的凸函数, 在点 $y_0 := F(x_0)$ 处取有限值, $\gamma \in \partial g(y_0)$ 满足

$$\langle\gamma, DF(x_0)h\rangle = g_-^{\downarrow}(y_0, DF(x_0)h). \tag{3.190}$$

设 g 在 y_0 处沿 $d := DF(x_0)h$ 处是外二阶正则的, 且 Robinson 约束规范 (3.87) 成立. 则复合函数 $f := g \circ F$ 在 x_0 处沿方向 h 对 $\alpha := [DF(x_0)]^*\gamma$ 是抛物正则的, 且

$$d^2 f(x_0\,|\alpha)(h) = \inf_{w\in X}\{g_-^{\downarrow\downarrow}(y_0; d, DF(x_0)w + D^2F(x_0)(h,h)) - \langle\alpha, w\rangle\}. \tag{3.191}$$

证明　根据命题 3.96, 由 g 的外二阶正则性可得, 复合函数 f 在 x_0 处沿 h 是外二阶正则的. 由于 $\gamma \in \partial g(y_0)$ 且 $F(x_0 + th') = F(x_0) + tDF(x_0)h' + o(t)$ 对 $h' \to h, t \downarrow 0$ 成立, 得到

$$g(F(x_0 + th')) - g(F(x_0)) - t\langle\gamma, DF(x_0)h'\rangle \geqslant o(t^2),$$

① 原著中上式为 $+\langle\alpha, w_n - w_n'\rangle$.

因而有 $d^2f(x\ |\alpha)(h)\geqslant 0$. 由命题 3.103 可得 f 的抛物正则性①. 公式 (3.191) 由 (3.187) 及相应的对复合函数的 (抛物) 下二阶上图导数的公式 (3.88) 得到. □

结合命题 3.100 与 3.101 的结论, 可以用 (抛物) 二阶下上图导数来描述二阶最优性条件.

命题 3.105 *设 x_0 是 (P) 的满足一阶必要条件 $f_-^{\downarrow}(x_0,\cdot)\geqslant 0$ 的可行点. 则*

(i) 若 x_0 是 (P) 的局部最优解, 则有

$$\inf_{w\in X} f_-^{\downarrow\downarrow}(x_0;h,w)\geqslant 0,\qquad \forall h\in C(x_0). \tag{3.192}$$

(ii) 若还有 X 是有限维空间, 且 f 在 x_0 处对 $\alpha=0$ 是抛物正则的, 则在 x_0 处二阶增长条件成立的充分必要条件是

$$\inf_{w\in X} f_-^{\downarrow\downarrow}(x_0;h,w)>0,\qquad \forall h\in C(x_0)\backslash\{0\}. \tag{3.193}$$

例 3.106 设 S 是 X 中的闭凸集, 考虑相应的指示函数$f(\cdot):=I_S(\cdot)$, 点 $x\in S$ 与方向 $h\in T_S(x)$. $f_-^{\downarrow}(x,h)=0, f_-^{\downarrow\downarrow}(x;h,\cdot)=I_{T_S^2(x,h)}(\cdot)$. 指示函数$f$ 在 x 处 (外) 二阶正则的充要条件是 S 在 x 处 (外) 二阶正则的. 令 $\alpha\in N_S(x)$ 满足 $\langle\alpha,h\rangle=0$. 由命题 3.103 可得, 若 S 在 x 处沿方向 h 是外二阶正则的, 则 f 在 x 处沿方向 h 对 α 是抛物正则的, 且

$$d^2f(x\ |\alpha)(h)=\inf_{w\in T_S^2(x,h)}(-\langle\alpha,w\rangle)=-\sigma(\alpha,T_S^2(x,h)). \tag{3.194}$$

有意思的是, 若 $\alpha=0$, 不管 S 是二阶正则的与否, 均有 $d^2f(x\ |\alpha)(h)=0$. 另一方面, $\inf_w f_-^{\downarrow\downarrow}(x;h,w)$ 等于 0 当且仅当 $T_S^2(x,h)$ 是非空的, 否则它是 $+\infty$. 所以, f 在 x 处沿方向 h 对 $\alpha=0$ 是抛物正则的充分必要条件是 $T_S^2(x,h)$ 是非空的.

这一节的结果可用于推导在 3.2 节与 3.3 节中讨论的约束问题的二阶最优性条件. 令 S 是 X 的闭子集, $f:X\to\mathbb{R}$ 是二次连续可微函数, 考虑问题

$$\min_{x\in S} f(x). \tag{3.195}$$

显然, 上述问题等价于增广实值函数 $\bar{f}(\cdot):=f(\cdot)+I_S(\cdot)$ 在 X 上的极小化问题.

令 $x_0\in S$ 是上述问题的可行点. 对 $h\in X$, 有

$$\bar{f}_-^{\downarrow}(x_0,h)=Df(x_0)h+I_{T_S(x_0)}(h).$$

则若 x_0 是 $f(\cdot)$ 在 S 上的局部极小点, 对所有的 $h\in T_S(x_0)$ 有 $Df(x_0)h\geqslant 0$. 进一步, 还有

$$\bar{f}_-^{\downarrow\downarrow}(x_0;h,w)=Df(x_0)w+D^2f(x_0)(h,h)+I_{T_S^2(x_0,h)}(w),$$

① 原著中为 g.

对任何 $h \in T_S(x_0)$ 与 $w \in X$ 成立, 从而有

$$\inf_{w \in X} \bar{f}_-^{\downarrow\downarrow}(x_0; h, w) = D^2 f(x_0)(h,h) - \sigma(-Df(x_0), T_S^2(x_0,h)).$$

由命题 3.105(i) 得, 若 x_0 是 $f(\cdot)$ 在 S 上的局部极小点, 则

$$D^2 f(x_0)(h,h) - \sigma(-Df(x_0), T_S^2(x_0,h)) \geqslant 0, \quad \forall h \in C(x_0). \tag{3.196}$$

不管 S 是否为凸集, 条件 (3.196) 成立. 函数 $\bar{f}$ 在 x_0 处是外二阶正则的当且仅当 S 在 x_0 处为外二阶正则的. 所以, 由命题 3.105(ii) 得, 若 X 是有限维的, 集合 S 在 x_0 处是外二阶正则的, 且 x_0 满足一阶必要最优条件, 则在 x_0 的二阶增长条件成立的充要条件是

$$D^2 f(x_0)(h,h) - \sigma(-Df(x_0), T_S^2(x_0,h)) > 0, \quad \forall h \in C(x_0)\backslash\{0\}. \tag{3.197}$$

现在设集合 S 由 $S := G^{-1}(K)$ 给出, 其中 K 是 Banach 空间 Y 的闭凸集, $G: X \to Y$ 是二阶连续可微的映射. 进一步, 设 Robinson 约束规范在点 x_0 处成立. 则由链式法则 (3.60), 有

$$T_S^2(x_0,h) = DG(x_0)^{-1}[T_K^2(G(x_0), DG(x_0)h) - D^2G(x_0)(h,h)]. \tag{3.198}$$

所以, 此种情况, 二阶必要性条件 (3.196) 等价于 (3.99)(或等价地, 等价于问题 (3.101) 最优值的非负性). 进一步, 由命题 3.88 得, 若 K 在 $G(x_0)$ 处是外二阶正则的, 则 S 在 x_0 处是外二阶正则的. 此种情形, 二阶充分条件 (3.197) 等价于二阶充分条件 (3.163).

3.4 具体结构

这一节, 我们讨论这一章前面的结果对某些特殊问题和结构的应用与推广. 即要讨论复合优化、精确罚函数、二次规划及原问题可简化的一些情形.

3.4.1 复合最优化

这一节讨论复合最优化的一阶与二阶最优性条件. 复合最优化即光滑映射与凸函数复合的极小化问题, 是另外一种研究最优性条件的方式. 设 X 与 Y 是 Banach 空间. 考虑最优化问题

$$\text{(P)} \qquad \min_{x \in X} \; g(F(x)), \tag{3.199}$$

其中 $g : Y \to \overline{\mathbb{R}}$ 是 l.s.c. 凸函数, $F : X \to Y$ 是光滑映射. 上述问题, 作为复合最优化问题等价于

$$\min_{(x,c)\in X\times\mathbb{R}} c \quad \text{s.t.} \quad (F(x), c) \in \text{epi } g. \tag{3.200}$$

所以通过取 $f(x,c) = c$, $G(x,c) := (F(x), c)$ 与 $K := \text{epi } g$, 上述问题可视为 (3.1) 定义的集合约束问题的特殊性况.

相反的结果也真. 即可把问题 (3.1) 写成下述形式

$$\min_{x\in X} f(x) + I_Q(x) + I_K(G(x)). \tag{3.201}$$

取

$$F(x) := (f(x), x, G(x)) : X \to \mathbb{R} \times X \times Y$$

与 $g(c,x,y) := c + I_Q(x) + I_K(y)$, 问题 (3.201) 具有形式 (3.199). 所以集合约束与复合优化两种形式事实上是等价的. 3.4.2 节进一步讨论如何将形式为 (3.1) 的集合约束问题简化为无约束最优化问题. 本节的余下部分将前几节发展起来的结果翻译为复合最优化的语言.

将问题 (P) 写为 (3.200) 的形式, 定义 $G(x,c) := (F(x), c), K := \text{epi } g$. 这一节, 设映射 F 是连续可微的. 称 $x \in X$ 是问题 (P) 的可行点, 若 $F(x) \in \text{dom} g$, 即 $g(F(x))$ 是有限的.

设 x_0 是 (P) 的可行点, 置 $c_0 := g(F(x_0)), y_0 = F(x_0)$. 因为

$$DG(x_0, c_0)(h, c) = (DF(x_0)h, c),$$

于是, 在点 (x_0, c_0) 处的 Robinson 约束规范 (2.178) 可以写为

$$(0,0) \in \text{int}\{(y_0, c_0) + DF(x_0)X \times \mathbb{R} - \text{epi } g\}.$$

很清楚, 上述条件等价于

$$0 \in \text{int}\{y_0 + DF(x_0)X - \text{dom } g\}. \tag{3.202}$$

注意到, 若 $g(\cdot)$ 在 y_0 处是连续的, 则 y_0 是 $\text{dom } g$ 的一个内部点, 因而 (3.202) 成立. 与约束规范 (3.202) 等价的各种条件的讨论, 见 2.3.4 节.

问题 (P) 的 Lagrange 函数是

$$L((x,c),(\lambda,\gamma)) = \langle \lambda, F(x)\rangle + c(1+\gamma).$$

我们观察到, 若 $\gamma \neq -1$, Lagrange 函数关于 c 取下确界的值为 $-\infty$, 且有

$$\sigma((\lambda,-1), \text{epi } g) = \sup\{\langle \lambda, x\rangle - c; (x,c) \in \text{epi } g\} = g^*(\lambda),$$

其中 $g^*(\cdot)$ 是 $g(\cdot)$ 的共轭函数. 所以 (P) 的 (共轭) 对偶是 (见 (2.298))

$$\text{(D)} \qquad \max_{\lambda\in Y^*}\{-g^*(\lambda)+\inf_{x\in X}\langle\lambda,F(x)\rangle\}, \tag{3.203}$$

其中常数 $\gamma=-1$ 被省略掉.

问题 (3.199) 可视为 2.5.3 节定义的 (2.291) 问题的特殊情况, 其中 $f(\cdot)=0$. 像在 (2.295) 那样将 (3.199) 参数化, (3.199) 的共轭对偶成为 (3.203).

注意, 集合 $K:=\text{epi}\ g$ 可定义为约束形式

$$K=\{(y,c)\in Y\times\mathbb{R}:h(y,c)\leqslant 0\},$$

其中 $h(y,c):=g(y)-c$. 因为 g 被设为是正常的, 因而它的定义域是非空的, 对函数 $h(y,c)$, Slater 条件总是成立的. 因此, 由命题 2.58, $K:=\text{epi}\ g$ 在点 (y_0,c_0) 处的切锥为

$$T_K(y_0,c_0)=\{(d,c):g^{\downarrow}(y_0,d)\leqslant c\}. \tag{3.204}$$

于是一阶必要条件 (3.16) 可表示为

$$[DF(x_0)]^*\lambda=0,\quad \gamma=-1,\quad (\lambda,\gamma)\in N_K(y_0,c_0). \tag{3.205}$$

(3.205) 中的最后条件意味着

$$\langle\lambda,d\rangle+\gamma c\leqslant 0,\quad \forall(d,c)\in T_K(y_0,c_0),$$

由于 $\gamma=-1$, 根据 (3.204), 上式可写为

$$\langle\lambda,d\rangle\leqslant g^{\downarrow}(y_0,d),\quad \forall d\in Y. \tag{3.206}$$

进一步, 条件 (3.206) 等价于条件 $\lambda\in\partial g(y_0)$(见命题 2.126(ii)). 所以, 一阶必要条件具有下述形式

$$[DF(x_0)]^*\lambda=0,\quad \lambda\in\partial g(y_0). \tag{3.207}$$

注意到, 在 (3.207) 中省掉了平凡条件 $\gamma=-1$. 满足上述一阶必要条件的 Lagrange 乘子的存在意味着 $\partial g(y_0)$ 是非空的, 即 $g(\cdot)$ 在 y_0 处是次可微的.

由定理 3.9 得到下述结果.

命题 3.107 *设 x_0 是 (3.199) 的局部最优解, 设约束规范 (3.202) 成立. 则满足最优性条件 (3.207) 的 Lagrange 乘子集 $\Lambda(x_0)$ 是 Y^* 的非空的凸的有界的弱 * 紧致子集.*

上述一阶最优性条件也可以通过直接推导而得到. 考虑函数 $\phi(x):=g(F(x))$. 由复合函数的链式法则 (2.256) 得, 在约束规范 (3.202) 下, $\phi(\cdot)$ 在 x_0 处是方向上图可微的, 且

$$\phi^{\downarrow}(x_0,h)=g^{\downarrow}(y_0,DF(x_0)h). \tag{3.208}$$

由命题 3.99(i) 得 $\phi^{\downarrow}(x_0,\cdot)\geqslant 0$ 是可行点 x_0 的局部最优性的必要条件. 结果, 若 x_0 是 (3.199) 的局部最优解, 约束规范 (3.202) 成立, 则

$$g^{\downarrow}(y_0, DF(x_0)h)\geqslant 0,\quad \forall h\in X. \tag{3.209}$$

进一步, 由关于次微分的相应的链式法则 (2.318) 用于 $\psi(\cdot):=\phi^{\downarrow}(x_0,\cdot)$, 可得到 $\partial\psi(0)=[DF(x_0)]^*(\partial g(y_0))$. 所以, 在约束规范 (3.202) 下, 条件 (3.207) 与 (3.209) 等价意味着 $0\in\partial\psi(0)$.

讨论由 (3.199) 或 (3.200) 给出的问题 (P) 的基于广义 Lagrange 乘子 (见 (3.22)) 的一阶必要条件. 有 $(0,\lambda)\in\mathbb{R}\times Y^*$ 是奇异的 Lagrange 乘子当且仅当

$$[DF(x_0)]^*\lambda=0,\quad \gamma=0,\quad (\gamma,x)\in N_K(y_0,c_0). \tag{3.210}$$

由于 (3.204), (3.210) 中的最后条件意味着 $\lambda\in\Gamma^-$, 其中 $\Gamma:=\operatorname{dom} g^{\downarrow}(x_0,\cdot)$. 注意到, 由于 $g^{\downarrow}(y_0,\cdot)$ 是凸的正齐次的, Γ 是凸锥. 还有 $\mathcal{R}_{\operatorname{dom} g}(y_0)\subset\Gamma\subset T_{\operatorname{dom} g}(y_0)$, 因此 $\Gamma^-=N_{\operatorname{dom} g}(y_0)$. 所以, 条件 (3.210) 可以写成下述形式

$$[DF(x_0)]^*\lambda=0,\quad \lambda\in N_{\operatorname{dom} g}(y_0). \tag{3.211}$$

然而, 由于 g 在 y_0 处是次可微的, 则由命题 2.126(iii) 有 $N_{\operatorname{dom} g}(y_0)$ 与 $\partial g(x_0)$ 的回收锥相同. 所以, 此种情况, 可把 (3.211) 写为

$$[DF(x^*)]^*\lambda=0,\quad \lambda\in[\partial g(y_0)]^{\infty}. \tag{3.212}$$

由命题 2.97, Robinson 约束规范 (3.202) 可推出 (如果 Y 是有限维的, 还是等价的) 条件

$$[DF(x_0)X]^{\perp}\cap N_{\operatorname{dom} g}(y_0)=\{0\}. \tag{3.213}$$

显然, 若 (3.213) 成立, λ 满足 (3.211), 则 $\lambda=0$. 因此, 在 Robinson 约束规范 (3.202) 下, 不存在奇异的广义的 Lagrange 乘子, 这当然与一般的理论是一致的.

考虑 (3.200) 形式的问题 (P), 设它是定义 2.163 意义下的凸问题. 这意味着映射 $G(x,c):=(F(x),c)$ 关于集合 $(-\operatorname{epi} g)$ 是凸的. 由命题 2.162 得, 这等价于定义 2.161 意义下的问题 (P) 的凸性.

对应的正则性条件 (2.312) 有下述形式

$$0\in\operatorname{int}\{F(X)-\operatorname{dom} g\}, \tag{3.214}$$

最优性条件 (3.8) 可以写为

$$x_0\in\underset{x\in X}{\operatorname{argmin}}\langle\lambda,F(x)\rangle,\quad \lambda\in\partial g(y_0). \tag{3.215}$$

由定理 3.6 得, 若 (P) 是凸的, x_0 是 (P) 的最优解且条件 (3.214) 成立, 则满足 (3.215) 的 Lagrange 乘子的集合是 Y^* 的非空的凸的弱 * 的紧致的子集.

二阶最优性条件

现在讨论由 (3.199) 或等价的 (3.200) 给出的问题 (P) 的二阶必要与二阶充分条件. 在本节的后半部分, 设映射 $F(x)$ 是二次连续可微的且 $g(y)$ 是下半连续的正常的凸函数.

以下计算在相应的二阶最优性条件中出现的 "sigma 项". 令 x_0 是 (P) 的可行点, $y_0 := F(x_0), c_0 := g(y_0)$, 令 $d \in Y$ 满足 $g^{\downarrow}(y_0, d)$ 是有限的. 由命题 3.30 中公式 (3.53) 可得

$$T^2_{\operatorname{epi} g}((y_0, c_0), (d, g^{\downarrow}(y_0, d))) = \{(w, \gamma) : g^{\downarrow\downarrow}_{-}(y_0; d, w) \leqslant \gamma\}. \tag{3.216}$$

置 $\mathcal{T} := T^2_{\operatorname{epi} g}((y_0, c_0), (d, g^{\downarrow}(y_0, d))), \psi(\cdot) := g^{\downarrow\downarrow}_{-}(y_0; d, \cdot)$. 则对 $\lambda \in \Lambda(x_0)$, 相应的 sigma 项为

$$\begin{aligned} \sigma((\lambda, -1), \mathcal{T}) &= \sup_{w,\gamma}\{\langle \lambda, w\rangle - \gamma : \psi(w) \leqslant \gamma\} \\ &= \sup_{w} \ \{\langle \lambda, w\rangle - \psi(w)\} = \psi^*(\lambda), \end{aligned} \tag{3.217}$$

其中 ψ^* 是函数 ψ 的共轭函数.

此时的临界锥可以写为下述形式

$$C(x_0, c_0) = \{(h, c) : g^{\downarrow}(y_0, DF(x_0)h) \leqslant 0, c \leqslant 0\} \tag{3.218}$$

(临界锥的定义见 (3.94)[①], 并回顾切锥 $T_K(y_0, c_0)$ 公式 (3.204)). 省略 c, 可将临界锥写为

$$\mathcal{C}(x_0) = \{h \in X : g^{\downarrow}(y_0, DF(x_0)h) \leqslant 0\}. \tag{3.219}$$

在约束规范 (3.202) 下, 由一阶必要条件 (见 (3.209)) 得 $g^{\downarrow}(y_0, DF(x_0)h) \geqslant 0, \forall h \in X$. 因此, 此种情况之下, 临界锥可以写为

$$\mathcal{C}(x_0) = \{h \in X : g^{\downarrow}(y_0, DF(x_0)h) = 0\}. \tag{3.220}$$

注意到, 由于 $g^{\downarrow}(y_0, \cdot)$ 是下半连续的凸函数, 锥 $\mathcal{C}(x_0)$ 是闭的凸的.

上述的计算连同定理 3.45 中给出的二阶必要条件可以推出下述结果.

定理 3.108 (二阶必要条件) 设 x_0 是问题 (P) 的局部最优解. 设映射 $F(x)$ 是二阶连续可微的, 函数 $g(y)$ 是 l.s.c. 正常的凸函数, 且约束规范 (3.202) 成立. 则对任何 $h \in \mathcal{C}(x_0)$ 及任意凸函数 $\phi(\cdot) \geqslant g^{\downarrow\downarrow}_{-}(y_0; DF(x_0)h, \cdot)$, 下述不等式成立

$$\sup_{\lambda \in \Lambda(x_0)} \{\langle \lambda, D^2_{xx}F(x_0)(h, h)\rangle - \phi^*(\lambda)\} \geqslant 0. \tag{3.221}$$

① 应该是 (3.95).

注意到, 由于在约束规范 (3.202) 下, 对任意的 $h \in \mathcal{C}(x_0)$, $g^{\downarrow}(y_0, DF(x_0)h)$ 是零, 因而是有限的, 上述定理中相应的下二阶方向上图导数是有定义的.

为了写出定理 3.86 形式的二阶充分条件, 需要计算对广义 Lagrange 乘子的 sigma 项. 设 $g^{\downarrow}(y_0, d)$ 是有限的, λ 是满足条件 (3.211) 的奇异的Lagrange 乘子. 则类似于 (3.217) 并用 (3.216), 得到

$$\begin{aligned}\sigma((\lambda,0),\mathcal{T}) &= \sup_{w,\gamma}\{\langle\lambda,w\rangle : \psi(w)\leqslant\gamma\}\\ &= \sup_{w}\ \{\langle\lambda,w\rangle : w\in\operatorname{dom}\psi\}\\ &= \sigma(\lambda,\operatorname{dom} g_{-}^{\downarrow\downarrow}(y_0;d,\cdot)).\end{aligned}\tag{3.222}$$

结果, 定理 3.86 形式的二阶充分条件可表述为下述形式.

定理 3.109 (二阶充分性条件) 设 X 是有限维空间, x_0 是 (P) 的可行点, 相联系的广义 Lagrange 乘子集合非空. 设映射 $F(x)$ 是二阶连续可微的, 函数 $g(y)$ 是 l.s.c. 正常的凸函数, 集合 $K := \operatorname{epi} g$ 在 (y_0, c_0) 处是外二阶正则的, 且还有下述条件成立: 对每一 $h \in \mathcal{C}(x_0)\backslash\{0\}$, 或者存在 (3.207) 的 Lagrange 乘子 λ, 满足对 $\psi(\cdot) := g_{-}^{\downarrow\downarrow}(y_0; DF(x_0)h, \cdot)$ 有下述不等式成立

$$\langle\lambda, D_{xx}^2 F(x_0)(h,h)\rangle - \psi^*(\lambda) > 0,\tag{3.223}$$

或 $g^{\downarrow}(y_0; DF(x_0)h) > -\infty$, 存在满足 (3.211) 的奇异 Lagrange 乘子λ 满足

$$\langle\lambda, D_{xx}^2 F(x_0)(h,h)\rangle - \sigma(\lambda, \operatorname{dom}\psi) > 0.\tag{3.224}$$

则存在 $\alpha > 0$, 对 x_0 的邻域中的所有的 x 有

$$g(F(x)) \geqslant g(F(x_0)) + \alpha\|x - x_0\|^2.\tag{3.225}$$

由二阶增长条件 (3.225) 当然可得到 x_0 是 (P) 的局部最优解.

注意到, 由 (3.219) 得, 若 $g^{\downarrow}(y_0, DF(x_0)h) > -\infty$ 对某一 $h \in \mathcal{C}(x_0)$ 成立, $g^{\downarrow}(y_0, DF(x_0)h)$ 是有限的, 则上述定理中的相应的函数 $\psi(\cdot)$ 是有定义的. 若 g 在 y_0 处是正则的, 则对任何 $d \in Y$, 有 $g^{\downarrow}(y_0, d) > -\infty$.

3.4.2 精确罚函数与增广对偶性

考虑最优化问题

$$\text{(P)}\qquad \min_{x\in X} f(x) \quad \text{s. t.} \quad G(x)\in K,\tag{3.226}$$

其中 K 是 Y 得非闭凸子集, $f : X \to \mathbb{R}, G : X \to Y$, X 与 Y 是 Banach 空间. 这一节讨论问题 (P) 的如下形式的惩罚函数

$$\theta_r(x) := f(x) + r\operatorname{dist}(G(x), K), \quad r > 0.\tag{3.227}$$

观察到, 若 $x \in \Phi := G^{-1}(K)$, 则 $\theta_r(x) = f(x)$. 因此, 若存在 $r \in \mathbb{R}$, $x_0 \in \Phi$ 是 $\theta_r(x)$ 的局部极小值点, 则 x_0 是 (P) 的局部最优解. 我们证明, 在某些情况下, 相反的结论也是正确的, 即若 x_0 是约束问题 (P) 的局部最优解, 则存在 $r > 0$ 满足 x_0 是 $\theta_r(x)$ 的无约束的局部极小点. 这种情况下, 称 $\theta_r(\cdot)$ 是问题 (P) 在 x_0 处的精确惩罚函数.

显然, 若对某一 $r > 0$, $\theta_r(\cdot)$ 是精确惩罚函数, 则对任何 $r' \geqslant r$, $\theta_{r'}(\cdot)$ 是精确的惩罚函数. 所以, 我们对参数 r 下界的估计是感兴趣的.

注意到, $\theta_r(x)$ 的 (无约束) 极小化问题可视为复合最优化问题 (3.199), 若定义 $F(x) := (f(x), G(x))$, $g(c, y) := c + r\ \mathrm{dist}(y, K)$, 它可表述为

$$\min_{(x,\gamma)\in X\times\mathbb{R}} \gamma \quad \text{s.t. } (f(x), G(x), \gamma) \in \mathrm{epi}\, g. \tag{3.228}$$

考虑距离函数 $d(y) := \mathrm{dist}(y, K)$. 这是连续的 (事实上甚至是模为 1 的 Lipschitz 连续的) 且是凸 (因为 K 是凸的) 函数. 结果 $g(c, y)$ 是连续的凸函数. 与约束问题 (P) 的 (3.201) 形式比较, 在 (3.201) 中由指示函数给出了 “无穷大惩罚” (infinite penalization), 而 $\theta_r(x)$ 中的惩罚项是有限值的且连续的. 由于 $g(c, y)$ 是连续的凸函数, 且处处次可微的. 对 $y_0 \in K$(见例 2.130) 有

$$\partial d(y_0) = N_K(y_0) \cap B_{Y^*}. \tag{3.229}$$

现在设函数 f 是凸连续的, 约束问题 (P) 是凸的 (见定义 2.163). 则问题 (3.228) 亦是凸的. 因为 g 是连续的, 问题 (3.228) 相应的约束规范总是成立的. 注意到, $\partial g(c, y) = \{1\} \times r\partial d(y)$, 由于在这一节设函数 f 是实值的, 其定义域与整个空间 X 重合, 所以, 由 (3.8)(也见 (3.215)) 给出的一阶最优性条件是充分必要的, 具有下述形式: $x_0 \in \Phi$ 是 $\theta_r(x)$ 的极小点当且仅当存在 $\lambda \in Y^*$ 满足

$$x_0 \in \operatorname*{argmin}_{x\in X} L(x, \lambda) \text{ 且 } \lambda \in r\partial d(y_0), \tag{3.230}$$

其中 $y_0 := G(x_0), L(x, \lambda) := f(x) + \langle \lambda, G(x) \rangle$ 是 (P) 的 Lagrange 函数. 由 (3.229), 条件 (3.230) 与约束问题 (P) 的相应的一阶最优性条件 (3.8) 的区别仅在于 (3.230) 的 Lagrange 乘子 λ 是否有小于或等于 r 的范数.

在凸的情形, 这导致精确罚函数与满足最优性条件 (3.8) 的问题 (P) 的 Lagrange 乘子集合 Λ_0 间的下述的简单的关系.

定理 3.110　设函数 f 是凸连续的, 约束问题 (P) 是凸的, x_0 是 (P) 的最优解, 则下述成立:

(i) 若 $\lambda \in \Lambda_0$ 是 (P) 的 Lagrange 乘子且 $r \geqslant \|\lambda\|$, 则 $\theta_r(\cdot)$ 是问题 (P) 在 x_0 处的精确惩罚函数.

(ii) 相反地, 若存在 $r>0$, $\theta_r(\cdot)$ 是 (P) 在 x_0 的精确惩罚函数, 则存在 $\lambda\in\Lambda_0$ 满足 $r\geqslant\|\lambda\|$.

证明 由于问题 (P) 的最优解设为是 x_0, 有 x_0 是可行的, $y_0:=G(x_0)\in K$. 设 λ 是 (P) 的 Lagrange 乘子且 $r\geqslant\|\lambda\|$. 由 (3.229) 有 $\lambda\in r\partial d(y_0)$, 因此最优性条件 (3.230) 成立. 这就得到 x_0 是 $\theta_r(\cdot)$ 的极小点, 因而 $\theta_r(\cdot)$ 是 (P) 在 x_0 处的精确惩罚函数.

相反地, 设 $\theta_r(\cdot)$ 是 (P) 在 x_0 处的精确惩罚函数. 则存在 λ 满足 (3.230). 这就得到 λ 是 (P) 的 Lagrange 乘子且 $r\geqslant\|\lambda\|$. 这就完成了证明. □

考虑下述条件: 存在 $c>0$ 及 x_0 是邻域内的所有的 x, 不等式

$$\operatorname{dist}(x,G^{-1}(K))\leqslant c\operatorname{dist}(G(x),K) \tag{3.231}$$

成立. 注意到, G 在 x_0 处关于 K 的度量正则性(G 的度量正则性的定义见 (2.165)) 可推出 (3.231). 尤其, 若映射 G 是连续可微的, 由 Robinson 约束规范可得到条件 (3.231)(见定理 2.87).

下面的结果表明, 条件 (3.231), 可视为度量正则性条件 (2.165) 为较弱的形式, 是精确惩罚函数存在的充分条件.

命题 3.111 设 x_0 是约束问题 (P) 的局部最优解, 条件 (3.231) 成立, f 在 x_0 的附近是 Lipschitz 连续的. 则对充分大的 r, $\theta_r(\cdot)$ 是 (P) 在 x_0 处的精确惩罚函数.

证明 考虑点 $x\in X$. 若 x 是 (P) 的可行点, 即 $x\in\Phi:=G^{-1}(K)$, 则由于 x_0 是 (P) 的局部最优解, 对于 x 充分接近 x_0, 有 $f(x)\geqslant f(x_0)$, 得到 $\theta_r(x)\geqslant\theta_r(x_0)$. 现在设 $x\notin\Phi$, 令 $x'\in\Phi$ 满足 $\|x-x'\|\leqslant(1+\varepsilon)\operatorname{dist}(x,\Phi)$ 对某一 $\varepsilon>0$ 成立. 注意到可行集 Φ 是非空的, 因为 $x_0\in\Phi$, 这样的点 x' 存在且当 $x\to x_0$ 时 x' 趋于 x_0. 因为 f 在 x_0 的附近是 Lipschitz 连续的, 设其模为 κ, 对充分接近 x_0 的 x, 有

$$\begin{aligned} f(x)\geqslant f(x')-\kappa\|x-x'\| &\geqslant f(x_0)-\kappa\|x-x'\| \\ &\geqslant f(x_0)-\kappa(1+\varepsilon)\operatorname{dist}(x,\Phi). \end{aligned}$$

结合 (3.231) 得到, 对任何 $r>c\kappa,\theta_r(\cdot)$ 是问题 (P) 在 x_0 处的精确惩罚函数, 其中 c 是 (3.231) 中的常数. □

上述结果表明, 精确惩罚性在较一般的假设下是可以得到的, 但没有给出参数 r 的准确估计 (sharp estimate). 定理 3.110 研究的凸情形, 该估计应该与 Lagrange 乘子的范数相联系. 若 f 与 G 是光滑的, r 的下述界估计是可以证明的.

定理 3.112 设 x_0 是 (P) 的局部最优解, f 与 G 是 Gâteaux 可微的, 且对某一 $r>0,\theta_r(\cdot)$ 是 (P) 在 x_0 点处的精确惩罚函数. 则存在 (P) 的 Lagrange 乘子 λ 满足一阶必要条件 (3.16), 且 $r\geqslant\|\lambda\|$.

证明　由于 f 与 G 是 Gâteaux 可微的, 由链式法则 (见命题 2.47), 由 (3.229) 得 $\theta_r(\cdot)$ 在 x_0 处是方向可微的, 且

$$\theta_r'(x_0, h) = Df(x_0)h + r\ \text{dist}(DG(x_0)h, T_K(G(x_0))). \tag{3.232}$$

所以, 若 x_0 是 $\theta_r(\cdot)$ 的局部极小点, 则 $h = 0$ 是函数

$$\phi_r(h) := Df(x_0)h + r\ \text{dist}(DG(x_0)h, T_K(G(x_0)))$$

的极小点. 函数 $\phi_r(h)$ 是下述问题的罚函数

$$\min_{h \in X} Df(x_0)h \quad \text{s.t.} \quad DG(x_0)h \in T_K(G(x_0)). \tag{3.233}$$

问题 (3.233) 是凸的, 由定理 3.110(ii) 得, 问题 (3.233) 在 $h = 0$ 处存在 Lagrange 乘子 λ, 满足 $r \geqslant \|\lambda\|$. 还有 λ 是问题 (P) 在 x_0 处的 Lagrange 乘子, 因此完成证明. 注意到上述的推证没有基于问题 (P) 的任何形式的约束规范. □

若 f 与 G 是 Gâteaux 可微的, 则 x_0 是 $\theta_r(\cdot)$ 的局部极小点的一阶必要条件可写为

$$D_x L(x_0, \lambda) = 0, \quad \lambda \in N_K(G(x_0)) \text{ 且 } \|\lambda\| \leqslant r. \tag{3.234}$$

注意, 除 (2.234) 中额外的条件 $\|\lambda\| \leqslant r$ 外, 上述条件与约束问题 (P) 的一阶必要条件相同.

若问题 (P) 不是凸的, 则一般来说, (3.234) 不是充分的, 为了确保 x_0 的局部极小性, 二阶条件是需要的. 设满足一阶必要条件 (3.16) 的约束问题的 Lagrange 乘子集合 $\Lambda(x_0)$ 是非空的, 考虑带有相应常数 $M > 0$ 的二阶充分条件 (3.145). 下述结果表明, 若 $M < r$, 则这些条件成为问题 (3.228) 的充分条件.

定理 3.113　设 x_0 是约束问题 (P) 的可行点. 设 f 与 G 是二阶连续可微的, 且对 $r > 0, 0 < M < r$, 二阶充分条件 (3.145)(对约束问题 (P)) 是成立的. 则存在 $\alpha > 0$, 对 x_0 的邻域中的所有的 x, 二阶增长条件成立

$$\theta_r(x) \geqslant \theta_r(x_0) + \alpha\|x - x_0\|^2, \tag{3.235}$$

因而 $\theta_r(\cdot)$ 是 (P) 在 x_0 处的精确惩罚函数.

证明　令 $\Lambda_M(x_0) := \{\lambda \in \Lambda(x_0) : \|\lambda\| \leqslant M\}$. 则对所有的 $\lambda \in \Lambda_M(x_0)$, 下述不等式成立

$$\theta_r(x) - L(x, \lambda) \geqslant (r - \|\lambda\|)\,\text{dist}(G(x), K) - \langle \lambda, G(x_0) \rangle. \tag{3.236}$$

事实上, 令 $\varepsilon > 0$ 及 $y \in K$ 满足 $\text{dist}(G(x), K) \geqslant \|G(x) - y\| - \varepsilon$. 用 $\langle \lambda, y - G(x_0) \rangle \leqslant 0$, 可得

$$\begin{aligned}\theta_r(x) - L(x,\lambda) &= r\,\mathrm{dist}(G(x),K) - \langle\lambda, G(x)\rangle \\ &\geqslant r\|G(x)-y\| - r\varepsilon - \langle\lambda, G(x)-y\rangle - \langle\lambda, y\rangle \\ &\geqslant (r-\|\lambda\|)\|G(x)-y\| - r\varepsilon - \langle\lambda, G(x_0)\rangle.\end{aligned}$$

可以取 ε 任意的小, 不等式 (3.236) 成立.

用反证法. 设定理的结论不真. 则存在序列 $x_n \in G^{-1}(K)$, 具有形式 $x_n = x_0 + t_n h_n, \|h_n\| = 1, t_n \downarrow 0$ 满足

$$\theta_r(x_n) \leqslant \theta_r(x_0) + o(t_n^2). \tag{3.237}$$

这表明 $f(x_n) \leqslant f(x_0) + o(t_n^2)$ 与 $\mathrm{dist}(G(x_n),K) = o(t_n^2)$, 因此有 $\mathrm{dist}(DG(x_0)h_n, T_K(G(x_0))) = o(1)$. 对充分大的 n, h_n 属于临界锥 $C_\eta(x_0)$. 用 (3.236), 得

$$\theta_r(x_n) \geqslant L(x_n,\lambda) - \langle\lambda, G(x_0)\rangle = \theta_r(x_0) + \frac{1}{2}D^2_{xx}L(x_0,\lambda)(h_n,h_n) + o(t_n^2).$$

由二阶充分性条件 (3.145), 在有界集 $\Lambda_M(x_0)$ 上取上确界 ($o(t_n^2)/t_n^2 \to 0$ 在此集上是一致的), 由上述不等式可推出

$$\theta_r(x_n) \geqslant \theta_r(x_0) + \frac{1}{2}\beta t_n^2 + o(t_n^2),$$

对 $\beta > 0$ 成立. 这与 (3.237) 是矛盾的, 从而完成了证明. □

考虑问题 (P) 的标准右端扰动, 即

$$(\mathrm{P}_y) \qquad \min_x f(x) \quad \text{s.t.} \quad G(x) + y \in K$$

相应的最优值函数$v(y) := \mathrm{val}(\mathrm{P}_y)$. 我们证明, 精确惩罚的性质与问题 (P) 的平静性是密切相关的.

定义 3.114 *称问题 (P) 是弱平静的 (weakly calm), 若存在 $r \geqslant 0$ 满足函数 $v(\cdot) + r\|\cdot\|$ 在 $y = 0$ 处取到极小. 此种情况也称问题 (P) 是以模 r 为弱平静的.*

由命题 2.148 得, 若问题 (P) 是平静的 (即 $v(u)$ 在 $u = 0$ 处是次可微的), 则它是弱平静的, 且若 $v(u)$ 是凸的, 则问题 (P) 是弱平静的当且仅当它是平静的.

定理 3.115 *设 x_0 是问题 (P) 的最优解. 则 $\theta_r(x)$ 在 X 上于 x_0 处取极小点的充分必要条件是 (P) 是以模 r 为弱平静的.*

证明 有

$$\begin{aligned}\inf_{y\in Y}\{v(y) + r\|y\|\} &= \inf_{y\in Y}\ \inf_{x\in X}\{f(x) + r\|y\| : G(x) + y \in K\} \\ &= \inf_{x\in X}\ \inf_{y\in Y}\{f(x) + r\|y\| : G(x) + y \in K\} \\ &= \inf_{x\in X}\ \{f(x) + r\ \mathrm{dist}(G(x),K)\} \\ &= \inf_{x\in X}\ \theta_r(x).\end{aligned}$$

结论得证. □

通过限制问题 (P) 的可行集为局部极小点 x_0 的邻域或约束的扰动集的邻域, 可以给出上述结果的局部形式. 现在来叙述这样的结果.

定理 3.116　设 x_0 是问题 (P) 的最优解. 则存在 $M>0$, 满足 $\theta_r(x)$ 在集合 $\{x\in X:\operatorname{dist}(G(x),K)<M\}$ 上在 x_0 处取到极小的充分与必要条件是函数 $v(y)+r\|y\|$ 在 $y=0$ 处达到局部极小点.

证明　令 $M>0$ 满足, 当 $\|y\|<M$ 时, 有 $v(0)\leqslant v(y)+r\|y\|$. 则

$$\begin{aligned}\inf_{\|y\|<M}\{v(y)+r\|y\|\}&=\inf_{\|y\|<M}\ \inf_{x\in X}\{f(x)+r\|y\|:G(x)+y\in K\}\\&=\inf_{x\in X}\ \inf_{\|y\|<M}\{f(x)+r\|y\|:G(x)+y\in K\}\\&=\inf_{x\in X;\ \operatorname{dist}(G(x),K)<M}\{f(x)+r\operatorname{dist}(G(x),K)\}\\&=\inf_{x\in X;\ \operatorname{dist}(G(x),K)<M}\theta_r(x).\end{aligned}$$

因为 $v(0)=f(x_0)=\theta_r(x_0)$, 这一下确界等于 $v(0)$ 当且仅当 θ_r 在集合 $\{x\in X:\operatorname{dist}(G(x),K)<M\}$ 在 x_0 处达到极小点. 这证得结论. □

现在考虑增广对偶方法. 为简便起见, 仅给出全局的描述, 局部描述也容易得到. 令 $v(y)$ 是问题 (不一定是凸的)(P_y) 的最优值函数, 对 $c\geqslant 0$, 考虑函数

$$v_c(y):=v(y)+c\|y\|^2.$$

称 λ 是 (P) 的增广 Lagrange 乘子, 若 $\lambda\in\partial v_c(0)$ 对 $c\geqslant 0$ 成立. 若问题 (P) 是凸的, 因而 $v(\cdot)$ 是凸的, 则对任何 $c\geqslant 0$,$\partial v(0)=\partial v_c(0)$, 因而由对偶定理, 增广 Lagrange 乘子λ 是 (P) 的参数对偶的最优解.

定理 3.117　设 x_0 是 (P) 的最优解, λ 是增广的 Lagrange 乘子. 则只要 $r>\|\lambda\|$, 就有 $\theta_r(\cdot)$ 是在 x_0 处的精确惩罚函数.

证明　令 $\varepsilon>0$. 只需注意到, 当 $\|y\|\leqslant\varepsilon/c$ 与 $r>\|\lambda\|+\varepsilon$, 有

$$v(y)-\langle\lambda,y\rangle+c\|y\|^2\leqslant v(y)-\langle\lambda,y\rangle+\varepsilon\|y\|\leqslant v(y)+r\|y\|.\tag{3.238}$$

条件 $\lambda\in\partial v_c(0)$ 意味着 (3.238) 的左端在 $y=0$ 处取得最小值 $v(0)$. 由于 (3.238) 的右端在 $y=0$ 处取相同的值, 它在 $y=0$ 处达到极小. 由定理 3.116 可得结论. □

注意, 若 Y 是与其对偶相同的 Hilbert 空间, 则

$$\begin{aligned}\inf_{y\in Y}\{v(y)-\langle\lambda,y\rangle+c\|y\|^2\}&=\inf_{y\in Y}\ \inf_{x\in X}\{f(x)-\langle\lambda,y\rangle+c\|y\|^2:G(x)+y\in K\}\\&=\inf_{x\in X}\ \inf_{y\in Y}\{f(x)+c\|y-(2c)^{-1}\lambda\|^2-(4c)^{-1}\|\lambda\|^2:\\&\qquad G(x)+y\in K\}\\&=\inf_{x\in X}\{f(x)+c\operatorname{dist}(G(x)+(2c)^{-1}\lambda,K)^2\\&\qquad-(4c)^{-1}\|\lambda\|^2\}.\end{aligned}$$

因此, λ 是增广 Lagrange 乘子当且仅当增广 Lagrange 函数

$$L_c(x,\lambda) := f(x) + c\,\mathrm{dist}(G(x) + (2c)^{-1}\lambda, K)^2 - (4c)^{-1}\|\lambda\|^2$$

满足 val(P)= $\inf\limits_{x\in X} L_c(x,\lambda)$. 尤其, 此种情况, 增广 Lagrange 函数在 (P) 中的每一最优解处取到极小值.

3.4.3 线性约束与二次规划

具有有限多个线性约束的问题需要特别的讨论, 因为这样的问题, 其最优性条件不需要约束规范成立, 对于线性的凹的或二次的目标函数, 具体的结果是可以得到的. 称问题 (P) 是线性约束的, 若它具有下述形式

$$\text{(P)} \qquad \min_{x\in X} f(x) \quad \text{s. t.} \quad Ax - b \in K, \tag{3.239}$$

其中 $A : X \to Y$ 是连续的线性映射, $K \subset Y$ 是闭凸多面体锥, $b \in Y$, 函数 $f(x)$ 是连续的. 若还有目标函数 $f(x)$ 是线性的, 则上述优化问题成为 2.5.7 节讨论的线性规划问题 (注意到, 为记号方便起见, 同 2.5.7 节讨论的相应的线性规划问题相比较, 在 (3.239) 中 b 的符号为 $-b$).

多面体锥 K 可表示为下述形式

$$K = \{y \in Y : \langle y_i^*, y\rangle \leqslant 0, i = 1, \cdots, p\},$$

其中 $y_i^* \in Y^*, i = 1, \cdots, p$. 则问题 (3.239) 中的可行集 Φ 可以写为下述形式

$$\Phi = \{x : \langle a_i, x\rangle \leqslant b_i, i = 1, \cdots, p\}, \tag{3.240}$$

其中 $a_i := A^* y_i^* \in X^*, b_i := \langle y_i^*, b\rangle \in \mathbb{R}, i = 1, \cdots, p$. 因此, 上述问题也可以写为下述的形式

$$\text{(P)} \qquad \min_{x\in X} f(x) \quad \text{s. t.} \quad \langle a_i, x\rangle \leqslant b_i, \quad i = 1, \cdots, p. \tag{3.241}$$

在本节的后面, 我们用线性约束问题 (P) 的上述 (3.241) 形式. 相应的 Lagrange 函数是

$$L(x,\lambda) := f(x) + \sum_{i=1}^{p} \lambda_i(\langle a_i, x\rangle - b_i).$$

对线性约束问题可以不需要约束规范, 在局部最优解 x_0 处推导出相应的一阶必要性条件. 事实上, 由于集合 K 是多面体集, 雷达锥 $\mathcal{R}_K(y)$ 是多面体锥且对所有的 $y \in K$, 均有它与 $T_K(y)$ 重合. 结果得到

$$T_\Phi(x_0) = \{h \in X : Ah \in T_K(y_0)\} = \{h \in X : \langle a_i, h\rangle \leqslant 0, i \in I(x_0)\},$$

其中 $y_0 := Ax_0 - b$, $I(x)$ 记在 x 处起作用的约束集合, 即

$$I(x) := \{i : \langle a_i, x\rangle = b_i, i = 1, \cdots, p\}. \tag{3.242}$$

由引理 3.7 可得, 若 f 是 Gâteaux 可微的, 则 $h = 0$ 是下述线性化问题的最优解

$$\min_{h\in X} Df(x_0)h \quad \text{s.t.} \ \langle a_i, h\rangle \leqslant 0, \quad i \in I(x_0). \tag{3.243}$$

线性化问题 (3.243) 是具有有限 (即 0) 的最优值的线性规划问题. 由定理 2.202 可得, (3.243) 的对偶问题的最优解集是非空的. 我们得到下述结果.

命题 3.118 设 (P) 是线性约束问题, 其目标函数 f 是 Gâteaux 可微的, 设 x_0 是 (P) 的局部最优解. 则存在 Lagrange 乘子 $\lambda \in \mathbb{R}^p$ 满足下述条件

$$D_x L(x_0, \lambda) = 0, \quad \lambda_i \geqslant 0, \quad \lambda_i(\langle a_i, x_0\rangle - b_i) = 0, \quad i = 1, \cdots, p. \tag{3.244}$$

注意到, 具有形式 $\langle a, x\rangle = b$ 的约束可分解成两个不等式约束 $\langle a, x\rangle \leqslant b$, $\langle a, x\rangle \geqslant b$. 所以, 我们给出的理论对有限多个等式与不等式约束的问题也是有效的. 注意到, 在将等式分裂成两个不等式的变换下, 约束规范并不是不变的, 对需要约束规范的非线性约束问题, 这一简化不再是有效的.

问题 (P) 在可行点 x_0 处的临界方向锥是

$$C(x_0) := \{h \in X : Df(x_0)h \leqslant 0, \langle a_i, h\rangle \leqslant 0, \ i \in I(x_0)\}, \tag{3.245}$$

点 x_0 是可行的, 称为是稳定的 (stationary), 若满足一阶条件 (3.244) 的 Lagrange 乘子集合 $\Lambda(x_0)$ 是非空的. 此种情况, $h = 0$ 是线性化问题的最优解, 在临界锥 $C(x_0)$ 的定义中, $Df(x_0)h \leqslant 0$ 可被等式 $Df(x_0)h = 0$ 代替 (见命题 3.10), 因此, $C(x_0)$ 即是线性化问题的最优解集.

我们来讨论二阶最优性条件, 先从二阶必要条件开始.

定理 3.119 设 $x_0 \in X$ 是 (P) 的局部最优解. 若 f 在 x_0 处是二阶 Gâteaux 可微的, 则对每一 $h \in C(x_0)$, 有条件 $Df(x_0)h = 0$ 与 $D^2 f(x_0)(h, h) \geqslant 0$ 成立.

证明 设 x_0 是 (P) 的局部最优解, h 是临界方向. 则对 $t > 0$ 充分小, $x_0 + th$[①] 是 (P) 的可行点. 由于 x_0 是 (P) 的局部最优解, 由 f 沿 h 的一阶展开式可得 $Df(x_0)h \geqslant 0$. 由于 h 是临界方向, 则 $Df(x_0)h = 0$. 进一步, 由

$$\lim_{t\downarrow 0} \frac{f(x_0 + th) - f(x_0)}{\frac{1}{2}t^2} = D^2 f(x_0)(h, h) \tag{3.246}$$

及 x_0 的局部最优性得上述等式左端是非负的. 结论得证. □

① 原著中是 $x + th$.

为了导出二阶充分条件, 很自然要在 Hilbert 空间 X 上讨论问题 (见例 3.66 后面的讨论). 我们叙述二阶充分性条件并把它们同二阶增长条件联系起来.

定理 3.120 设 X 是 Hilbert 空间, x_0 是 (P) 的稳定点. 设 f 在 x_0 处是二阶连续可微的. 则在 x_0 处的二阶增长条件成立的充分必要条件是存在 $\alpha > 0$, 满足

$$D^2 f(x_0)(h,h) \geqslant \alpha \|h\|^2, \quad \forall h \in C(x_0). \tag{3.247}$$

证明 若在 x_0 处二阶增长条件成立, 则存在某一 $\varepsilon > 0$, x_0 是下述问题的局部最优解

$$\min_{x \in \Phi} \{f(x) - \varepsilon \|x - x_0\|^2\}.$$

将定理 3.119 应用于上述优化问题, 即得必要性条件 (3.247).

现在证条件 (3.247) 是二阶增长条件的充分条件. 假设在 x_0 处二阶增长条件不成立, 则存在收敛到 x_0 的可行点序列 x_k, 满足

$$f(x_k) \leqslant f(x_0) + o(\|x_k - x_0\|^2). \tag{3.248}$$

置 $t_k := \|x_k - x_0\|, h_k := (x_k - x_0)/t_k$. 对 $i \in I(x_0)$ 及充分大的 k, 有 $\langle a_i, h_k \rangle \leqslant 0$ 成立. 所以, 由 Hoffman 引理(定理 2.200) 得

$$\operatorname{dist}(h_k, C(x_0)) = O([Df(x_0)h_k]_+). \tag{3.249}$$

进一步, 由 (3.248) 得 $t_k Df(x_0)h_k \leqslant o(t_k)$, 因而存在临界方向 $\hat{h}_k \in C(x_0)$ 满足 $\hat{h}_k - h_k \to 0$, 从而 $\|\hat{h}_k\| \to 1$. 由于 x_0 是稳定点, 即 $\Lambda(x_0) \neq \varnothing$, 有 $h = 0$ 是线性化问题的最优解, 因而对充分大的 k, 有 $Df(x_0)h_k \geqslant 0$. 结果, 由 (3.247) 有

$$\begin{aligned} f(x_k) &= f(x_0) + t_k Df(x_0)h_k + \frac{1}{2} t_k^2 D^2 f(x_0)(\hat{h}_k, \hat{h}_k) + o(t_k^2)) \\ &\geqslant f(x_0) + \frac{1}{2} t_k^2 \alpha \|\hat{h}_k\|^2 + o(t_k^2) \\ &= f(x_0) + \frac{1}{2} \alpha \|x_k - x_0\|^2 + o(t_k^2). \end{aligned} \tag{3.250}$$

这与 (3.248) 矛盾. □

现在讨论当目标函数是凹函数时最优解集的结构. 由定理 2.198, 线性约束的凹最优化问题 (定义为在广义多面体集上求凹费用函数的极小化问题, 这里仅处理多面体可行集合) 在其最优值是有限的情况下, 最优解集是非空的. 集合 $M \subset \{1, \cdots, p\}$ 联系着 (多面的) 可行集 Φ 的 (可能空的) 面 F_M(见 (2.383)), 定义为

$$F_M := \{x \in \Phi : \langle a_i, x \rangle = b_i, \forall i \in M\}.$$

注意到, Φ 的每一面是多面集且 Φ 是其面的并.

定理 3.121 设 (P) 是线性约束的凹问题. (P) 的最优解集是 Φ 的有限多个面的 (可能空的) 并集.

证明 若 (P) 的最优解集 $\mathcal{S}(\mathrm{P})$ 是空的, 结论显然成立. 因此, 可设 $\mathcal{S}(\mathrm{P}) \neq \varnothing$. 我们证下述关系式成立

$$\mathcal{S}(\mathrm{P}) = \bigcup_{x \in \mathcal{S}(P)} F_{I(x)}, \tag{3.251}$$

其中 $I(x) \subset \{1, \cdots, p\}$ 记在 x 处起作用的约束的集合. 注意到 Φ 的不同面的个数是有限的. 所以, 定理的结果由上述公式得到.

现在证 (3.251). 设 $x_0 \in \mathcal{S}(\mathrm{P})$ 及 $x \in F_{I(x_0)}$, 置 $h := x - x_0, x(t) := x_0 + th, t \in \mathbb{R}$. 注意到 $\langle a_i, x(t) \rangle = b_i, \forall i \in I(x_0)$ 及 $t \in \mathbb{R}$. 由连续性, 对 $i \in I(x_0)$ 及充分接近于 0 的 t 有 $\langle a_i, x(t) \rangle < b_i$. 因此, 对充分接近于零的 t, $x(t)$ 是可行的. 进一步, 有 $x(1) = x \in F_{I(x_0)}$, 因此存在 $t_0 > 0$ 满足只要 $t \in [-t_0, 1]$ 有 $x(t) \in \Phi$. 观察到, 若凹函数定义于区间 $[a, b] \subset \mathbb{R}$ 上, 其最小点在此区间的内部点上取到, 则函数在 $[a, b]$ 上是常数. 结果, $f(x(t))$ 在区间 $[-t_0, 1]$ 上取常值, 因而有 $f(x) = f(x_0)$, 即 $x \in \mathcal{S}(\mathrm{P})$. 证得 $F_{I(x_0)} \subset \mathcal{S}(\mathrm{P})$. 由于 $\mathcal{S}(\mathrm{P}) \subset \Phi$, Φ 是它的面的并, 从而公式 (3.251) 成立. □

例 3.122 考虑问题

$$\min_{x \in \mathbb{R}^2} f(x) \quad \text{s.t.} \ \ x_1 + x_2 \leqslant 1, \quad x_1 \geqslant 0, \quad x_2 \geqslant 0,$$

其中 $f(x) := \min\{x_1, x_2\}$. 这一问题的最优解集由连接 $(0,0)$ 与 $(1,0)$ 和连接 $(0,0)$ 与 $(0,1)$ 的这两个线段的并组成.

在上述例子中, 最优解集合有一个 “角”. 下述命题表明, 这种情况只可能当目标函数是不可微的时候发生.

命题 3.123 设 X 是 Banach 空间, 设函数 f 是凹的且是 Gâteaux 可微的. 则 $\mathcal{S}(\mathrm{P})$ 是集合 Φ 的有限多个互不相交的面的并. 即 $\mathcal{S}(\mathrm{P})$ 的连通部分是 Φ 的面.

证明 由定理 3.121, $\mathcal{S}(\mathrm{P})$ 是 Φ 的有限多个面的并. 用反证法. 假设结论是错误的. 则存在两个不同的面 $F_{M_1}, F_{M_2} \subset \mathcal{S}(\mathrm{P})$ 满足它们的交是非空的, 且不存在面 $F_M \subset \mathcal{S}(\mathrm{P})$ 包含这两个面. 则存在 x_0, x_1, x_2 满足 $x_0 \in F_{M_1} \bigcap F_{M_2}, x_1 \in F_{M_1} \backslash F_{M_2}, x_2 \in F_{M_2} \backslash F_{M_1}$ 且 $(x_1 + x_2)/2 \notin \mathcal{S}(\mathrm{P})$. 置 $h_1 := x_1 - x_0, h_2 := x_2 - x_0$. 则函数 f 在区间 $[x_0, x_1]$ 与 $[x_0, x_2]$ 上取常数值, 因而有 $Df(x_0)h_i := 0, i = 1, 2$. 因为 f 是可微的且是凹的, 则得到 $f\left(x_0 + \dfrac{1}{2}(h_1 + h_2)\right) \leqslant f(x_0)$. 显然有 $x_0 + \dfrac{1}{2}(h_1 + h_2) = (x_1 + x_2)/2 \in \Phi$. 这就得到 $(x_1 + x_2)/2 \in \mathcal{S}(\mathrm{P})$, 这导致矛盾. □

对于线性约束的凹的可微问题, 有下述 Lagrange 乘子的性质.

命题 3.124 设 $f: X \to \mathbb{R}$ 是凹的 Gâteaux 可微的, 则在 $\mathcal{S}(\mathrm{P})$ 的每一连通部分上 Lagrange 乘子集合是常值的.

证明 由命题 3.123, $\mathcal{S}(\mathrm{P})$ 的连通部分 F 是 Φ 的一个面, 因而它是凸的多面体集合. 由引理 2.119, $Df(x)$ 在凸集 F 上是常值的, 如 $Df(x) = \gamma$. 由线性化的推证有, F 是下述线性化问题

$$\min_{x \in \Phi} \langle \gamma, x \rangle$$

的最优解集的子集. 因为这是凸问题, 此线性化的Lagrange 乘子集合在 F 上是常值. 因为 $\gamma = Df(x)$, 该线性化问题的 Lagrange 乘子集合与原问题在 $x \in F$ 的 Lagrange 乘子集重合. □

现在讨论线性约束的凹问题的增长条件.

命题 3.125 若 Banach 空间中的线性规划问题具有非空的最优解集, 则它满足一阶增长条件.

证明 令 $f(x) = \langle \alpha, x \rangle$ 是 (P) 的目标函数. 有

$$\mathcal{S}(P) = \{x \in X, \langle a_i, x \rangle \leqslant b_i, i = 1, \cdots, p, \langle \alpha, x \rangle = \mathrm{val}(\mathrm{P})\},$$

即集合 $\mathcal{S}(\mathrm{P})$ 可由有限多个线性约束来定义. 由 Hoffman 引理(定理 2.200), 由于对任意的 $x \in \Phi$ 有 $\langle \alpha, x \rangle \geqslant \mathrm{val}(\mathrm{P})$, 存在常数 $k > 0$ 满足

$$\mathrm{val}(\mathrm{P}) + k^{-1} \operatorname{dist}(x, \mathcal{S}(\mathrm{P})) \leqslant \langle \alpha, x \rangle, \quad \forall x \in \Phi. \tag{3.252}$$

□

注意, 一阶增长条件 (3.252) 在下述意义下是 "全局的", 即对每一 $x \in \Phi$ 均成立 (而不是只限定在最优集 $\mathcal{S}(\mathrm{P})$ 的邻域).

注 3.126 令 (P) 是线性约束的凹问题. 若空间 X 的维数大于 p (尤其若 X 是无限维的,) 则 (P) 的最优解集 $\mathcal{S}(\mathrm{P})$ 或者是空集或者是无界的. 事实上, 令 x_0 是 (P) 的最优解. 则集合

$$S := \{x \in X : \langle a_i, x - x_0 \rangle = 0, i = 1, \cdots, p\}$$

是非空的 (因为 $x_0 \in S$), 它是 X 中的具有正维数的仿射子空间, 因而是无界的. 很清楚, S 的每一点均是 (P) 的可行点, 由于 f 是凹的, 有 $S \subset \mathcal{S}(\mathrm{P})$.

注意到, 若 x_0 是 (P) 的局部最优解, f 在 x_0 处是可微的, $h \in C(x_0)$ 是临界方向, 则 (由 f 的凹性) 对所有的 $t > 0$ 充分小, 有 $x_0 + th$ 也是 (P) 的局部最优解. 所以, 若 x_0 是 (P) 的孤立的局部最优解, 则 $C(x_0) = \{0\}$. 很显然, 只有 X 是有限维的, 这才是可能的. 在 X 是有限维的情况, 若 (P) 是凹的线性约束的, 函数 f 是可微的, 由在 3.1.4 给出的一阶充分条件的理论, 得到在 (P) 的每一孤立局部最优解 x_0 处一阶增长条件是成立的.

这一节的余下部分, 设空间 X 是 Hilbert 空间, 讨论下述形式的二次规划问题

$$\text{(QP)} \qquad \min_{x \in X} q(x) \quad \text{s.t.} \quad \langle a_i, x \rangle \leqslant b_i, \quad i = 1, \cdots, p, \tag{3.253}$$

其中

$$q(x) := \langle \alpha, x \rangle + \frac{1}{2} Q(x), \tag{3.254}$$

$\alpha \in X^*$, $Q(\cdot)$ 是二次型. 函数 $Q : X \to \mathbb{R}$ 称为是二次型, 若 $Q(x) = B(x, x)$, 其中 $B : X \times X \to \mathbb{R}$ 是双线性的对称函数. 在后续的分析中假设二次型 Q 是 Legendre 形式 (见 3.3.2), 即 Q 是弱下半连续的, 满足若 $x_k \xrightarrow{w} x$ 且 $Q(x_k) \to Q(x)$, 则 $x_k \to x_0$. 由命题 3.71 得, 二次型 Q 是凸的当且仅当它是非负的. 由命题 2.111 还可以得到, 若 Q 是凸的, 则它在弱拓扑下是下半连续的当且仅当它在强拓扑下是下半连续的. 因而得到, 非负的弱 l.s.c. 二次型是连续的 (在强拓扑下).

现在讨论二次规划问题的最优解的存在性, 先考虑等式约束 (可能是无限多个约束) 的问题

$$\text{(EQP)} \qquad \min_{x \in X} q(x) \quad \text{s.t.} \quad Ax = b, \tag{3.255}$$

其中 $q(x)$ 定义为 (3.254) 的形式, A 是由 X 到 Banach 空间 Y 的连续的线性算子, $b \in Y$. 设 $b \in \text{Range}(A)$, 即存在一点 $x_0 \in X, Ax_0 = b$. 则 (EQP) 等价于问题

$$\min_{h \in X} q(x_0 + h) \quad \text{s.t.} \quad Ah = 0. \tag{3.256}$$

用 $\mathcal{N}(Q) := \{x \in X : Q(x) = 0\}$ 记二次型 Q 的零空间, 用 $\mathcal{N}(A) = \ker A$ 记算子A 的零空间. 若二次型是非负的, 则 $\mathcal{N}(Q)$ 是线性空间, 若 $(x_1, x_2) \in \mathcal{N}(Q) \times X$, 有 $Q(x_1 + x_2) = Q(x_2)$(见命题 3.72).

引理 3.127　下述性质成立:

(i) 设 (EQP) 的最优解集 $\mathcal{S}$(EQP) 是非空的, $x_0 \in \mathcal{S}$(EPQ). 则二次型 Q 在 $\mathcal{N}(A)$ 上是非负的, 且

$$\mathcal{S}(\text{EPQ}) = x_0 + \mathcal{N}(A) \cap \mathcal{N}(Q).$$

(ii) 设 Q 是 Legendre 形式. 则下述条件是等价的: (a) 问题 (EQP) 具有最优解; (b) 问题 (EQP) 是可行的, 二次型 Q 在 $\mathcal{N}(A)$ 上是非负的, 对任何可行点 $\bar{x}$, 对所有的 $h \in \mathcal{N}(A) \cap \mathcal{N}(Q)$ 有 $Dq(\bar{x})h = 0$; (c) 二次型Q 在 $\mathcal{N}(A)$ 上是非负的, 且存在可行点 $\bar{x}$, 对所有的 $h \in \mathcal{N}(A) \cap \mathcal{N}(Q)$ 满足 $Dq(\bar{x})h = 0$.

(iii) 若 x_0 是 (EQP) 的局部最优解, 则 x_0 是 (EQP) 的全局最优解.

证明　(i) 验证是初等的, 留给读者. 下证 (ii). 令 $\bar{x}$ 是 (EQP) 的可行点, $h \in \mathcal{N}(A)$. 若 $Q(h) < 0$, 则只要 $t \to \pm\infty$, 有 $q(\bar{x} + th) \to -\infty$. 若 $Q(h) = 0, Dq(\bar{x})h \neq 0$,

如有必要, 不妨将 h 变化为 $-h$, 可设 $Dq(\bar{x})h < 0$. 则当 $t \to +\infty$ 时 $q(\bar{x}+th) \to -\infty$. 所以, 若 (b) 不成立, 则 (a) 不成立.

(b) 推出 (c) 是显然的, 剩下的需要证 (c) 可推出 (a). 这等价于证问题

$$(\mathrm{EQP}') \qquad \min_{h \in \mathcal{N}(A)} \quad q(\bar{x}+h)$$

至少有一解. 由命题 3.79, Q 到 Banach 空间 $\mathcal{N}(A)$ 上的限制 Q_1 亦是 Legendre 形式, 它是非负的, 因此 $\mathcal{N}(Q_1)$ 是 $\mathcal{N}(A)$ 的闭的向量子空间. 令 X_1 是它在 $\mathcal{N}(A)$ 中的正交补空间. 任何 $h \in \mathcal{N}(A)$ 可唯一地表示为 $h = h_1 + h_2$, 其中 $h_1 \in X_1, h_2 \in \mathcal{N}(Q_1)$. 这意味着

$$q(\bar{x}+h_1+h_2) = q(\bar{x}) + Dq(\bar{x})h_1 + Q(h_1) = q(\bar{x}+h_1).$$

由命题 3.80 有 Q_1 在 X_1 上是椭圆型的, 因此问题

$$\min_{h_1 \in X_1} \quad q(\bar{x}+h_1)$$

具有唯一的最优解 $\hat{x}$. 所以, (EQP$'$) 具有一解, 这完成 (ii) 的证明.

设 x_0 是 (EQP) 的局部最优解. 则有 $Q_2(h) := q(x_0+h) - q(x_0) : \mathcal{N}(A) \to \mathbb{R}$ 是二次型 (因为在 0 处的它的导数不会非零) 在 $h = 0$ 处取到局部极小. 这就得到 Q_2 是非负的, 因而是凸的, 从而 $h = 0$ 是 Q_2 的全局极小点. □

现在给出二次规划(QP) 的最优解的存在性的刻画, 它们是对线性约束的凹问题所得到结果的回顾.

定理 3.128 设二次型 Q 是 Legendre 形式. 则下述条件等价:

(i) 问题 (QP) 有最优解.

(ii) (QP) 的最优值有限.

(iii) 问题 (QP) 是相容的, 且对任何可行点 $\bar{x}$, 不存在 h 满足 $\langle a_i, h\rangle \leqslant 0, i = 1, \cdots, p$, 使得 $Q(h) < 0$ 或者 $Q(h) = 0$ 且 $Dq(\bar{x})h < 0$.

证明 递推关系 (i)⇒(ii) 是显然的. 设 (ii) 成立, 则 (QP) 是相容的, 因为否则有 val(QP)$= +\infty$. 若 $\bar{x}$ 是 (QP) 的可行点, 且存在 h 满足 $\langle a_i, h\rangle \leqslant 0, i = 1, \cdots, p, Q(h) < 0$ 或者 $Q(h) = 0$ 且 $Dq(\bar{x})h < 0$, 则 $t \to +\infty$ 时 $q(\bar{x}+th) \to -\infty$, 因而有 val(QP)$= -\infty$, 与 (ii) 矛盾. 这证得递推关系 (ii)⇒(iii).

现在证由 (iii) 可推出 (i). 由 (3.242) 定义的 $I(x)$, 记起作用约束的集合. 设 $\{x_k\}$ 是问题 (QP) 的极小化序列. 因为 $\{1, \cdots, p\}$ 可能的子集的个数是有限的, 若有必要可取一子列, 设 $I(x_k)$ 是常值集合, 如等于 I. 进一步, 以下面的方式选取极小化序列 $\{x_k\}$: 若 $\hat{x}_k$ 是另一极小化序列, 满足对所有的 k 均有 $I \subset I(\hat{x}_k)$, 则对所有的 k 均有 $I = I(\hat{x}_k)$, 即起作用约束集 $I(x_k)$ 是 "极大的". 我们来证明下述问题具有最优

解:

$$\min_{x \in X} q(x) \quad \text{s.t.} \quad \langle a_i, x \rangle = b_i, \quad i \in I. \tag{3.257}$$

若这不是真的, 则由引理 3.127, 存在 $h_k \in \mathcal{N}(A)$ 满足 $\langle a_i, h_k \rangle = 0, \forall i \in I$, 或者 $Q(h_k) < 0$ 或者 $Q(h_k) = 0$ 且 $DQ(x_k)h_k < 0$. 后一种情况, 若有必要将 h_k 换为 $-h_k$, 可不妨设 $Dq(\bar{x})h_k \leqslant 0$. 两种情况, 都有函数 $t \mapsto q(x_k + th_k) : \mathbb{R}_+ \to \mathbb{R}$ 当 $t \uparrow +\infty$ 时递减到 $-\infty$. 由假设, 存在某一 $j \in \{1, \cdots, p\} \backslash I$, 有 $\langle a_j, h_k \rangle > 0$, 令 $t_k \in (0, \infty)$ 是使 $\hat{x}_k := x_k + t_k h_k$ 为 (QP) 可行点的最大值. 因为 $q(\hat{x}_k) \leqslant q(x_k)$, 有 $\{\hat{x}_k\}$ 是极小化序列且 $I(\hat{x}_k)$ 严格包含 $I(x_k)$. 抽出常值的起作用约束的子列, 得到这与 $\{x_k\}$ 的构造是矛盾的. 这样证得 (3.257) 有最优解, 记为 $\bar{x}$.

现在证 $\bar{x}$ 是问题 (P) 的可行点. 假设这是错误的. 令 $\hat{x}_k$ 是区间 $[x_k, \bar{x}]$ 中的距离 $\bar{x}$ 最近的 (QP) 的可行点. 由于 $\hat{x}_k \in [x_k, \bar{x}]$, 有 $I \subset I(\hat{x}_k)$, 由构造知, 在 $\hat{x}_k$ 处至少有一起作用的额外的不等式约束存在, 因此 $I \neq I(\hat{x}_k)$. 再由问题 (3.257) 有最优解, 二次型$Q(\cdot)$ 在线性空间 $\{x : \langle a_i, x \rangle = 0, i \in I\}$ 上是非负的, 因而 $q(\cdot)$ 在 (3.527) 的可行集上是凸的, 因此在区间 $[x_k, \bar{x}]$ 上是凸的. 可得到

$$q(\hat{x}_k) \leqslant \max\{q(x_k), q(\bar{x})\} = q(x_k).$$

因而 $\{\hat{x}_k\}$ 是 (QP) 的极小化序列, $I \subset I(\hat{x}_k)$, 且 $I \neq I(\hat{x}_k)$. 这与集合 I 的 "极大性" 矛盾. 证得 $\hat{x}_k$ 是问题 (P) 的可行点.

最后, 由于 $\{x_k\}$ 是问题 (QP) 的极小化序列, 有 $\lim\limits_{k \to \infty} q(x_k) =$val(QP), 因为 x_k 是问题 (3.257) 的可行点, $q(\bar{x}) \leqslant \lim\limits_{k \to \infty} q(x_k)$. 从而 $q(\bar{x}) \leqslant$ val(QP). 由于 $\bar{x}$ 是 (QP) 的可行点, 有 $q(\bar{x}) \geqslant$ val(QP), 因此 $q(\bar{x}) =$val(QP). 于是 $\bar{x}$ 是 (QP) 的最优解. 这就证得递推关系 (iii)⇒(i). □

若空间 X 是有限维的, 则任何二次型 $Q : X \to \mathbb{R}$ 均是 Legendre 形式. 因此, 有限维二次规划问题有最优解的充分必要条件是其最优值是有限的.

再来讨论二次规划问题 (QP) 的最优解集的结构. 对集合 $M \subset \{1, \cdots, p\}$, 我们联系着下述等式约束的问题

$$(\text{QP}_M) \qquad \min_{x \in X} q(x) \quad \text{s.t.} \ \langle a_i, x \rangle = b_i, \quad i \in M. \tag{3.258}$$

定理 3.129 设 Φ 是二次规划问题 (QP) 的可行集, $\mathcal{S}(\text{QP}_M)$ 是 (QP_M) 的最优解集. 则

$$\text{val(QP)} = \min\{\text{val}(\text{QP}_M) : \mathcal{S}(\text{QP}_M) \cap \Phi \neq \varnothing, M \subset \{1, \cdots, p\}\}, \tag{3.259}$$

$$\mathcal{S}(\text{QP}) = \cup\{\mathcal{S}(\text{QP}_M) \cap \Phi : \text{val}(\text{QP}_M) = \text{val(QP)}, M \subset \{1, \cdots, p\}\}. \tag{3.260}$$

证明 令 $\bar{x}$ 在 $\mathcal{S}(\mathrm{QP})$ 中, 置 $\overline{M}:=I(\bar{x})$. 下证 $\bar{x}\in\mathcal{S}(\mathrm{QP}_{\overline{M}})$. 事实上, $\bar{x}$ 是 $(\mathrm{QP}_{\overline{M}})$ 的可行点, 限制到 $\bar{x}$ 的充分小邻域, $(\mathrm{QP}_{\overline{M}})$ 的可行集是 (QP) 的可行集的子集. 所以, $\bar{x}$ 是 $(\mathrm{QP}_{\overline{M}})$ 的局部最优解. 则由引理 3.127(iii) 得 $\bar{x}$ 是 $(\mathrm{QP}_{\overline{M}})$ 的全局最优解, 有 $\mathrm{val(QP)}=\mathrm{val}(\mathrm{QP}_{\overline{M}})$, 由此易得 (3.259) 与 (3.266). □

由于 $\mathcal{S}(\mathrm{QP}_M)\cap\Phi$ 是仿射空间与多面体集合的交, 它也是多面体集合. 结果, 由公式 (3.260) 得 $\mathcal{S}(\mathrm{QP})$ 是有限多个多面体集合的并集.

现在给出二次规划问题 (QP) 的一阶与二阶最优性条件. 一阶与二阶必要性条件是前面的结果的简单应用. 二阶充分性条件有些令人惊奇, 因为二阶必要条件刚好既是局部最优性的必要条件, 也是充分性条件.

定理 3.130 设 x_0 是问题 (QP) 的可行点, 则

(i) 若 x_0 是 (QP) 的局部最优解, 则相应的 Lagrange 乘子的集合非空.

(ii) 点 x_0 是 (QP) 的局部最优解的充分必要条件是相应的 Lagrange 乘子集合非空, 且对所有的 $h\in C(x_0)$, 有 $Q(h)\geqslant 0$.

证明 结论 (i) 是命题 3.118 的直接的结论. 我们证 (ii). 令 x_0 是 (QP) 的局部最优解. 有 $D^2q(x_0)(h,h)=Q(h)$, 因此, 由定理 3.119 有对所有的 $h\in C(x_0)$ 满足 $Q(h)\geqslant 0$.

现在证相反的推出关系 (将这一证明与定理 3.120 的证明相比较). 令 x_0 是 (QP) 的稳定点并设对所有的 $h\in C(x_0), Q(h)\geqslant 0$. 假设 x_0 不是 (QP) 的局部最优解. 则存在收敛于 x_0 的可行点列 x_k, 满足 $q(x_k)<q(x_0)$. 置 $t_k:=\|x_k-x_0\|$ 及 $h_k:=(x_k-x_0)/t_k$. 则像定理 3.120 的证明一样, 由 Hoffman 引理(定理 2.200), 存在临界方向 $\hat{h}_k\in C(x_0)$ 满足

$$\|\hat{h}_k-h_k\|=O([Dq(x_0)h_k]_+). \tag{3.261}$$

因为 $\Lambda(x_0)\neq\varnothing$, 有 $h=0$ 是线性化问题的最优解, 因而对充分大的 k, 有 $Dq(x_0)h_k\geqslant 0$. 结果, 由于对函数 $q(\cdot)$, 二阶 Taylor 展式是精确的, 由 (3.261) 得

$$\begin{aligned}q(x_k)&=q(x_0)+t_kDq(x_0)h_k+\frac{1}{2}t_k^2Q(h_k)\\&=q(x_0)+t_kDq(x_0)h_k+\frac{1}{2}t_k^2Q(\hat{h}_k)+t_k^2\eta_k,\end{aligned}$$

其中 $\eta_k=O([Dq(x_0)h_k]_+)$. 则对充分大的 k 有

$$q(x_k)-q(x_0)\geqslant\frac{1}{2}t_k^2Q(\hat{h}_k)\geqslant 0,$$

这与 $q(x_k)<q(x_0)$ 矛盾. □

我们讨论线性约束最优化问题的严格互补性.

定义 3.131 设 (P) 是形式为 (3.241) 的线性约束的最优化问题. 称 (P) 的可行点 x 与其相应的 Lagrange 乘子 $\lambda=(\lambda_1,\cdots,\lambda_p)\in\Lambda(x)$ 是严格互补的, 若对任何满足 $\langle a_i,x\rangle=b_i$ 的 $i\in\{1,\cdots,p\}$ 有 $\lambda_i>0$. 称问题 (P) 满足严格互补条件若存在最优解 x 及 Lagrange 乘子 $\lambda\in\Lambda(x)$, 它们是严格互补的.

我们证明, 对线性规划问题与线性约束的凹问题, 严格互补条件总是成立的. 下述例子说明, 对于凸的二次规划问题, 严格互补条件不总是成立的.

例 3.132 考虑问题

$$\min_{x\in\mathbb{R}} x^2 \quad \text{s.t.}\ \ x\geqslant 0.$$

此问题具有唯一最优解 $x_0=0$, 与之相联系的唯一的 Lagrange 乘子 $\lambda=0$. 显然, 严格互补条件并不成立.

对 (P) 的可行点 x 与相应的 Lagrange 乘子 $\lambda\in\Lambda(x)$, 定义指标集

$$\begin{aligned}
I_P(x)&:=\{i:\langle a_i,x\rangle-b_i<0,i=1,\cdots,p\},\\
I_D(\lambda)&:=\{i:\lambda_i>0,i=1,\cdots,p\},\\
I_R(x,\lambda)&:=\{1,\cdots,p\}\backslash(I_P(x)\cup I_D(\lambda)).
\end{aligned}\tag{3.262}$$

一阶最优性条件 (3.244) 中的互补条件推出 $I_P(x)\cap I_D(\lambda)=\varnothing$. 因此有 $(I_P(x)$, $I_D(\lambda)$, $I_R(x,\lambda))$ 是 $\{1,\cdots,p\}$ 的分划. 显然, x 与 λ 是严格互补的充要条件是 $I_R(x,\lambda)=\varnothing$. 在凸的情况, 即线性约束问题 (P) 中的目标函数是凸的情况, 有 Lagrange 乘子λ 是 (P) 的对偶问题的最优解 (见命题 3.3).

设问题 (P) 是凸的, 最优解集$\mathcal{S}(\mathrm{P})$ 也是凸的. 若 $x,x'\in\mathcal{S}(\mathrm{P})$, 则对任何在区间 $[x,x']$ 中的 x'', $I_P(x)\subset I_P(x'')$, $I_P(x')\subset I_P(x'')$. 因此 $I_P(\cdot)$ 在 $\mathcal{S}(\mathrm{P})$ 中的某一凸子集上具有最大的常值 I_P, 即对任何 $x\in\mathcal{S}(\mathrm{P})$, $I_P(x)\subset I_P$. 若 $\mathcal{S}(\mathrm{P})$ 是非空集合, 则这样的子集由它的相对内部给出. 类似地, $I_D(\lambda)$ 在 $\mathcal{S}(\mathrm{D})$ 的凸子集合上也取常值 I_D. 令 $I_R:=\{1,\cdots,p\}\backslash(I_P\cup I_D)$. 由于 $I_P\cap I_D$ 是空集, 有 (I_P,I_D,I_R) 是 $\{1,\cdots,p\}$ 的分划, 称为最优分划. 若 I_R 是空集, 称凸问题(P) 满足严格互补条件.

现在考虑 (3.241) 形式的线性规划问题, 其中 $f(x):=\langle\alpha,x\rangle,\alpha\in X^*$. 下述结果表明, 对线性规划问题而言, 严格互补条件总是成立的. 由命题 2.197 得, 任何非空多面体集合具有非空的相对内部.

定理 3.133 设 X 是 Banach 空间, 问题 (P) 是具有非空最优解集的线性规划问题. 则 (P) 满足严格互补条件.

证明 因为 $\mathcal{S}(\mathrm{P})$ 是非空的, 由定理 2.202 可得 $\mathcal{S}(\mathrm{D})$ 是非空集. 只需证明, 若 $x\in\operatorname{ri}\mathcal{S}(\mathrm{P})$, $\lambda\in\operatorname{ri}\mathcal{S}(\mathrm{D})$, 则 x 与 λ 是严格互补的, 即集合 $I_R=I_R(x,\lambda)$ 是空集. 用反证法. 设 I_R 是非空的, $j\in I_R$. 令 $I_P=I_P(x)$, $I_D=I_D(\lambda)$, 考虑如下线性规划问

题

$$(\mathrm{LP}_j) \qquad \min_{h\in X}\langle a_j,h\rangle \quad \text{s.t.} \quad \langle\alpha,h\rangle\leqslant 0, \quad \langle a_i,h\rangle\leqslant 0, \quad i\in I_D\cup I_R.$$

我们论断 $\mathrm{val}(\mathrm{LP}_j)=0$. 实际上, 由 $h=0$ 是 (PL_j) 的可行点, 得到 $\mathrm{val}(\mathrm{LP}_j)\leqslant 0$. 如果 $\mathrm{val}(\mathrm{LP}_j)<0$, 则存在 (PL_j) 的一可行点 $\bar h$ 满足 $\langle a_j,\bar h\rangle<0$. 由于对任何满足 $\langle a_i,x\rangle=b_i$ 的 $i=\{1,\cdots,p\}$ 有 $\langle a_i,\bar h\rangle\leqslant 0$, 则对 $\varepsilon>0$ 充分小 $x(\varepsilon):=x+\varepsilon\bar h$ 是问题 (P) 的可行点. 有 $\langle\alpha,x(\varepsilon)\rangle\leqslant\langle\alpha,x\rangle$, 因而对 $\varepsilon>0$ 充分小有 $x(\varepsilon)\in\mathcal{S}(\mathrm{P})$. 因为 $\langle a_j,x(\varepsilon)\rangle-b_j=\varepsilon\langle a_j,\bar h\rangle<0$, 这得到一与 I_R 的定义相矛盾的结果. 这就证得我们的论断.

问题 (LP_j) 的 Lagrange 函数是

$$\langle a_j,h\rangle+\sum_{i\in I_D\cup I_R}\lambda_i\langle a_i,h\rangle+\mu\langle\alpha,h\rangle,$$

其对偶问题可以写为

$$(\mathrm{DP}_j) \quad \max_{\lambda,\mu}\ 0 \quad \text{s.t.} \sum_{i\in I_D\cup I_R}\lambda_i a_i+\mu\alpha+a_j=0, \quad \lambda_i\geqslant 0, \quad i\in I_D\cup I_R, \quad \mu\geqslant 0.$$

因为 $\mathrm{val}(\mathrm{LP}_j)=0$, 对偶问题$(\mathrm{DP}_j)$ 至少具有一最优解 $(\bar\lambda,\bar\mu)$. 定义 $\lambda^*\in\mathbb{R}^p$:

$$\lambda_i^*:=\bar\lambda_i, \quad i\in I_D\cup I_R\backslash\{j\}, \quad \lambda_j^*:=\bar\lambda_j+1, \quad \lambda_i^*=0, \quad i\in I_P\ .$$

分两种情况. 若 $\bar\mu>0$, 则 $\hat\lambda:=\lambda^*/\bar\mu$ 满足

$$\alpha+\sum_{1\leqslant i\leqslant p}\hat\lambda_i a_i=0, \quad \hat\lambda_i\geqslant 0, \quad \hat\lambda_i(\langle a_i,x\rangle-b_i)=0, \quad i=1,\cdots,p.$$

得到 $\hat\lambda\in\mathcal{S}(\mathrm{D})$. 则 $\hat\lambda_j=(\bar\lambda_j+1)/\bar\mu>0$, 与 I_D 的定义矛盾. 若为另一种情况, $\bar\mu=0$, 在 $\mathcal{S}(\mathrm{D})$ 中取 λ, 置 $\hat\lambda:=\lambda+\lambda^*$. 则有 $\hat\lambda\in\mathcal{S}(\mathrm{D})$ 且 $\hat\lambda_j>0$, 这与 I_D 的定义矛盾. □

由命题 3.123 得, 若目标函数 f 是凹的且可微的, 则 $\mathcal{S}(\mathrm{P})$ 的每一连通部分是 Φ 的面, 在这一面上 f 的导数是常值.

定理 3.134 设 x 是具有连续可微目标函数的线性约束的凹问题 (P) 的局部最优解. 则对属于 $\mathcal{S}(\mathrm{P})$ 的连通部分的相对内部的每一问题 (P) 的最优解, 联系着满足严格互补条件的 Lagrange 乘子.

证明 设 F 是 $\mathcal{S}(\mathrm{P})$ 的连通部分, 令 c 是在 F 上的 Df 的公共值. 由线性化的论证, 可得 F 即是在 (多面集) Φ 上求 $\langle c,x\rangle$ 的极小化的线性规划问题的最优解集. 本定理中的结论由定理 3.133 的结果可以得到. □

3.4.4 一种简化的方式

在某些情况下, 集合 K 可以被重新参数化于被考虑的点 $y_0 \in K$ 的一邻域内, 使得原始问题 (P) 被简化为简单的问题.

这一节我们研究这样一种方式并证明在这样的情况下, sigma 项可解释为与光滑映射相联系的曲率. 这一节设 X, Y 与 Z 是 Banach 空间, K 是 Y 的闭凸子集, (P) 是如下的最优化问题

$$(\mathrm{P}) \qquad \min_{x \in X} f(x) \quad \text{s. t.} \quad G(x) \in K, \tag{3.263}$$

其中 f 与 G 是二次连续可微的, x_0 是 (P) 的可行点, 即 $y_0 := G(x_0) \in K$. 映射 $\Xi := Y \to Z$ 称为是 C^l 光滑的, $l \in \mathbb{N} \cup \{\infty\}$, 若 Ξ 是 l 次连续可微的. 锥 $\mathcal{C} \subset Z$ 称为是点的, 若 $z \in \mathcal{C}$ 与 $-z \in \mathcal{C}$ 可推出 $z = 0$. 包含在 $\mathcal{C}$ 中的最大的线性空间被称为 $\mathcal{C}$ 的线 (lineality) 空间. 当然, 锥 $\mathcal{C}$ 是点的当且仅当其线性空间是 $\{0\}$.

定义 3.135　*令 $K \subset Y, \mathcal{C} \subset Z$ 是凸闭集. 称集合 K 于点 $y_0 \in K$ 处是 C^l 简约到集合 $\mathcal{C}$ 的 (C^l-reducible to the set $\mathcal{C}$), 若存在 y_0 的邻域 N 及 l 阶连续可微的映射 $\Xi : N \to Z$ 满足*

(i) *$D\Xi(y_0) : Y \to Z$ 是映上的, 且*

(ii) *$K \cap N = \{y : \Xi(y) \in \mathcal{C}\}$.*

称这一简约是点的, 若切锥 $T_{\mathcal{C}}(\Xi(y_0))$ 是点锥. 若还有集合 $\mathcal{C} - \Xi(y_0)$ 是点的闭凸锥, 称 K 在 y_0 处是 C^l 锥简约的 (C^l-cone reducible). 不妨假设 $\Xi(y_0) = 0$.

当然, 上述定义中的映射 Ξ 的选取不是唯一的且是有点随意性的. 然而, 在几个重要的例子里, 这样的映射可以很自然地被构造出来. 本书中讨论的简约化方法的应用与例子中, 空间 X, Y 与 Z 通常是有限维的. 然而, 这一点并不是本质性的. 注意到, 若 $\mathcal{C} \subset Z$ 是凸锥, 空间 Z 是有限维的, 则 $\mathcal{C}$ 可表示为它的线空间与点的凸锥 $\mathcal{C}'$ 的直和. 通过取到由 $\mathcal{C}'$ 生成的子空间的直交投影, 我们可把此问题简化为相应的锥是点的情形. 所以, 不失一般性, 只讨论锥简化为点锥的情形.

上述定义中的条件 (ii) 意味着集合 K 可以局部地定义为约束 $\Xi(y) \in \mathcal{C}$, 因此, 在 x_0 的附近, (P) 的可行集可局部地由约束 $\mathcal{G}(x) \in \mathcal{C}$ 来定义, 其中 $\mathcal{G}(x) := \Xi(G(x))$. 结果, 在 x_0 的邻域, 原问题 (P) 等价于下面所谓的简化问题:

$$(\mathcal{P}) \qquad \min_{x \in X} f(x) \quad \text{s. t.} \quad \mathcal{G}(x) \in \mathcal{C}. \tag{3.264}$$

(P) 与 $(\mathcal{P})$ 的可行集在 x_0 附近是重合的, 因此, 限制在 x_0 的邻域上, (P) 与 $(\mathcal{P})$ 的最优解集是相同的.

设集合 K 在 y_0 处是 C^1 简约到集合 $\mathcal{C}$ 的, 则问题 $(\mathcal{P})$ 在点 x_0 处的一阶最优性条件可以写为

$$D_x \mathcal{L}(x_0, \mu) = 0, \quad \mu \in N_{\mathcal{C}}(\mathcal{G}(x_0)), \tag{3.265}$$

其中

$$\mathcal{L}(x,\mu) := f(x) + \langle \mu, \mathcal{G}(x) \rangle$$

是简化问题的 Lagrange 函数. 由条件 (i), $D\Xi(y_0)$ 是映上的 (即 $D\Xi(y_0)Y = \mathcal{Z}$), 对充分接近于 y_0 的所有的 y, $z := \Xi(y)$, 有

$$D\Xi(y)[T_K(y)] = T_{\mathcal{C}}(z).$$

有 $\mu \in N_{\mathcal{C}}(z)$ 当且仅当对任何 $h \in T_K(y)$,

$$0 \geqslant \langle \mu, D\Xi(y)h \rangle = \langle D\Xi(y)^* \mu, h \rangle.$$

这意味着

$$N_K(y) = D\Xi(y)^* N_{\mathcal{C}}(z). \tag{3.266}$$

于是, 若 $\Lambda(x_0)$ 与 $M(x_0)$ 分别是问题 (P) 与 $(\mathcal{P})$ 的 Lagrange 乘子的集合, 则

$$\Lambda(x_0) = [D\Xi(y_0)]^* M(x_0). \tag{3.267}$$

注意到, 由于 $D\Xi(y_0)$ 是映上的, 其伴随映射 $D\Xi(y_0)^*$ 是一对一的. 也注意到, 问题 (P) 的 Robinson 约束规范于 x_0 处成立当且仅当 $(\mathcal{P})$ 在 x_0 处的 Robinson 约束规范成立.

现在设 K 是在 y_0 处是 C^2 锥简约为集合 $\mathcal{C}$ 的. 因为 $\mathcal{C}$ 是凸锥, 对任何 $z \in \mathcal{C}$, 有

$$T_{\mathcal{C}}^2(0,z) = T_{\mathcal{C}}^{i,2}(0,z) = T_{\mathcal{C}}(z) = \mathrm{cl}\,\{\mathcal{C} + [\![z]\!]\}. \tag{3.268}$$

结果, $0 \in T_{\mathcal{C}}^{i,2}(0,z)$, $\mathcal{C}$ 在 $0 \in \mathcal{Z}$ 处是二阶正则的. 由于 $D\Xi(y_0)$ 是映上的, 由命题 3.88, 集合 K 在 y_0 处是二阶正则的, 尤其对任何方向 d 二阶内切集 $T_K^{i,2}(y_0,d)$ 与外二阶切集 $T_K^2(y_0,d)$ 是重合的. 进一步, 由二阶切集的链式法则 (见公式 (3.59)), 有

$$T_K^2(y_0,d) = D\Xi(y_0)^{-1}[T_{\mathcal{C}}^2(0, D\Xi(y_0)d) - D^2\Xi(y_0)(d,d)]. \tag{3.269}$$

得到下述结果.

命题 3.136 *设集合 K 于 y_0 处 C^2 锥简约到锥 $\mathcal{C}$, 则 K 在 y_0 处是二阶正则的, 且公式 (3.269) 成立.*

进一步, 由于 $\mathcal{C}$ 是凸锥, 对简化问题$(\mathcal{P})$, 其 sigma 项是零, 在 Robinson 约束规范下, 下述二阶条件对 x_0 的局部最优性是必要的 (见定理 3.45).

$$\sup_{\mu \in M(x_0)} D_{xx}^2 \mathcal{L}(x_0,\mu)(h,h) \geqslant 0, \quad \forall h \in C(x_0), \tag{3.270}$$

其中 $C(x_0)$ 是问题 $(\mathcal{P})$ 的临界锥. 由于

$$\begin{aligned} C(x_0) := &\{h \in X : D\mathcal{G}(x_0)h \in \mathcal{C}, Df(x_0)h \leqslant 0\} \\ = &\{h \in X : DG(x_0)h \in T_K(y_0), Df(x_0)h \leqslant 0\}, \end{aligned} \tag{3.271}$$

因此问题 (P) 与 $(\mathcal{P})$ 的临界锥是相同的.

由链式法则有

$$D^2_{xx}\mathcal{L}(x_0, \mu)(h, h) = D^2_{xx}L(x_0, \lambda)(h, h) + \varsigma(\mu, h), \tag{3.272}$$

其中 $\lambda := [D\Xi(y_0)]^*\mu$ 且

$$\varsigma(\mu, h) := \langle \mu, D^2\Xi(y_0)(DG(x_0)h, DG(x_0)h)\rangle. \tag{3.273}$$

由于 $\lambda := [D\Xi(y_0)]^*\mu$, 由 (3.268) 与 (3.269) 有

$$\sigma(\lambda, T^{i,2}_K(y_0, DG(x_0)h)) = -\varsigma(\mu, h), \tag{3.274}$$

此种情况下, (3.272) 中出现的额外的项 $\varsigma(\mu, h)$ 正好是问题 (P) 的 "sigma 项". 所以, 由 (3.272) 与 (3.274) 有 (3.270) 的二阶条件与定理 3.45 中的二阶必要条件是相同的.

定理 3.63 中的二阶充分条件也可用于简化的问题. 这些条件不含有 sigma 项 (对简化的问题, 它是零), 不需要 X 的有限维数的性质. 对原始问题 (P), 翻译成相应的条件, 它们 (在 Robinson 约束规范下) 取下述形式

$$\sup_{\lambda \in \Lambda(x_0)} \{D^2_{xx}L(x_0, \lambda)(h, h) - \sigma(\lambda, \mathcal{T}(h))\} \geqslant \beta\|h\|^2, \quad \forall h \in \widehat{C}_\eta(x_0), \tag{3.275}$$

其中

$$\mathcal{T}(h) := T^{i,2}_K(y_0, DG(x_0)h),$$

β 与 η 是正常数, $\widehat{C}_\eta(x_0)$ 由 (3.147) 定义. 注意到, 由 (3.273), 上述二阶条件的 sigma 项是 h 的二次函数. 进一步, 若空间 X 是有限维的, 则上述二阶条件可以写为下述形式

$$\sup_{\lambda \in \Lambda(x_0)} \{D^2_{xx}L(x_0, \lambda)(h, h) - \sigma(\lambda, \mathcal{T}(h))\} > 0, \quad \forall h \in C(x_0)\backslash\{0\}. \tag{3.276}$$

得到下述结果.

定理 3.137 设空间 X 是有限维的. 令 x_0 是满足一阶最优性条件的 (P) 的可行点. 设集合 K 在 y_0 处是 C^2 锥简约的, 且 Robinson 约束规范在 x_0 点成立. 则 (3.276) 是在点 x_0 处的二阶增长条件的充分与必要条件.

考虑一些例子. 若 Z 是有限维的, 如 $Z := \mathbb{R}^k$, 则映射 $\Xi(y) = (\xi_1(y), \cdots, \xi_k(y))$ 定义为坐标的局部系统. 此种情形 "$D\Xi(y_0)$ 是映上的" 意味着 $D\xi_1(y_0), \cdots, D\xi_k(y_0)$ 是线性无关的.

例 3.138 令 K 在 y_0 的附近由有限多个不等式约束定义. 即存在 C^l 光滑的函数 $\xi_1(y), \cdots, \xi_k(y)$ 及 y_0 的邻域 N 满足 $\xi_i(y_0) = 0, i = 1, \cdots, k$, 且

$$K \cap N = \{y \in N : \xi_1(y) \geqslant 0, \cdots, \xi_k(y) \geqslant 0\}. \tag{3.277}$$

进一步设 $D\xi_1(y_0), \cdots, D\xi_k(y_0)$ 是线性无关的. 此种情况, 集合 K 在 y_0 处 C^l 锥简约为锥 $\mathcal{C} := \mathbb{R}^k_+$ 的.

例 3.139 令 K 是有限维空间的多面体子集. 考虑切锥 $T_K(y_0)$. 因为 K 是多面体集合, 则存在 y_0 的邻域 N 满足在此邻域上 K 与 $y_0 + T_K(y_0)$ 相同. 令 $L := \mathrm{lin}[T_K(y_0)]$ 是 $T_K(y_0)$ 的线空间, 即 $T_K(y_0)$ 的最大的线性子空间, 则锥 $T_K(y_0)$ 由有限多个线性约束来定义, $T_K(y_0) = \{y : a_i^{\mathrm{T}} y \geqslant 0, i = i, \cdots, p\}$, 其中 $a_i \in L^\perp$, $L^\perp$ 是与 L 垂直的 (正交的) 线性空间且满足 $L + L^\perp = Y$. 令 $k := \dim(L^\perp)$, 取 $L^\perp$ 的一组基 $b_1, \cdots, b_k$. 则 $b_1, \cdots, b_k$ 是线性无关的, 每一向量 $a_i, i = 1, \cdots, p$ 可表示为向量 $b_1, \cdots, b_k$ 的线性组合. 于是集合 K 可局部地由下述的坐标系统 $\xi_i(y) := b_i^{\mathrm{T}}(y - y_0), i = 1, \cdots, k$, 来定义. 在此坐标系统下集合 K 是 C^∞ 锥简约的, 其中 $\mathcal{C}$ 由锥 $\mathbb{R}^k_+$ 给出.

例 3.140 考虑半定规划的例子 (见例 2.65). 即令 $Y := \mathcal{S}^p$ 是 $p \times p$ 对称矩阵空间, $K := \mathcal{S}^p_+$ 是由 $p \times p$ 半正定矩阵构成的锥. 令 $A_0 \in \mathcal{S}^p_+$ 是秩为 $r < p$ 的矩阵. 用 $\lambda_1(A) \geqslant \cdots \geqslant \lambda_p(A)$ 表示 $p \times p$ 对称矩阵 A 的特征值, 用 $e_1(A), \cdots, e_p(A)$ 记相应的特征向量. 令 $E(A)$ 是以 $e_{r+1}(A), \cdots, e_p(A)$ 为列构成的 $p \times (p - r)$ 矩阵. 因为 $E(A)^{\mathrm{T}} A E(A)$ 是对角元素是 $\lambda_{r+1}(A), \cdots, \lambda_p(A)$ 的对角矩阵且特征值 $\lambda_1(A_0), \cdots, \lambda_r(A_0)$ 是正的, 在 A_0 的邻域上, 锥 $\mathcal{S}^p_+$ 可由约束 $E(A)^{\mathrm{T}} A E(A) \succeq 0$ 来定义 (记号 $B \succeq 0$ 意味着矩阵 B 是半正定的). 后一约束在 $(p-r) \times (p-r)$ 对称矩阵空间内取值. 然而, 这一表示不适合于我们的目的, 因为特征向量 $e_{r+1}(A), \cdots, e_p(A)$ 不是唯一定义的, 除非 0 是 A_0 的简单特征值, 当 A 靠近 A_0, $E(A)$ 不是连续的 (因而不是光滑的) 函数. 为克服这一困难, 有下述做法 (下述构造类似于例 3.98 中的构造).

记 $L(A)$ 为对应于 A 的 $p - r$ 个最小特征值的特征空间, $P(A)$ 是到 $L(A)$ 上的直交投影矩阵. 令 E_0 是 (固定的)$p \times (p - r)$ 矩阵, 它的列是直交的且张成空间 $L(A_0)$, 即 $E_0 := E(A_0)$. 已知在 A_0 的充分小的邻域内, $P(A)$ 是 A 的连续可微的 (其实是解析的) 函数. 结果, $F(A) := P(A)E_0$ 也是 A_0 的邻域内 A 的连续可微函数, 进一步, $F(A_0) = E_0$. 对充分靠近 A_0 的所有的 A, $F(A)$ 的秩是 $p - r$, 即它的列向量是线性无关的. 令 $U(A)$ 是用 Gram-Schmidt 正交化过程作用于 $F(A)$ 的列生成的列

向量的矩阵. 则矩阵 $U(A)$ 是有定义的且在 A_0 附近是连续可微的, 进一步满足下述条件: $U(A_0)=E_0, U(A)$ 的列空间与 $E(A)$ 的列空间重合, 且 $U(A)^{\mathrm{T}}U(A)=I_{p-r}$.

我们得到在 A_0 的邻域 N 内, 锥 $\mathcal{S}_+^p$ 可定义为

$$\{A\in\mathcal{S}^p:\ U(A)^{\mathrm{T}}AU(A)\succeq 0\}. \tag{3.278}$$

考虑映射 $\Xi: A\to U(A)^{\mathrm{T}}AU(A)$, 由 N 映到 $\mathcal{S}^{p-r}$. 映射 Ξ 是连续可微的, 由于 $D\Xi(A_0)A$ 的表达式是

$$(DU(A_0)^{\mathrm{T}}A)A_0U(A_0)+U(A_0)^{\mathrm{T}}AU(A_0)+U(A_0)^{\mathrm{T}}A_0(DU(A_0)A),$$

由 $A_0U(A_0)=0$, 有 $D\Xi(A_0)A=E_0^{\mathrm{T}}AE_0$. 得到 $D\Xi(A_0)$ 是映上的. 于是 $\mathcal{S}_+^p$ 在 A_0 处 C^∞ 锥简约为锥 $\mathcal{C}:=\mathcal{S}_+^{p-r}$.

由命题 3.136 得, 对所有的 $p\in\mathbb{N}$, $p\times p$ 正半定矩阵的锥在每一点处均是二阶正则的.

3.5 非孤立的极小点

这一节讨论具有非孤立最优解集的最优化问题的二阶最优性条件. 例如, 考虑凸优化问题. 因为凸问题的最优解是凸的, 若最优解不是唯一的, 则它有非孤立的最优解. 当然, 在凸的情况, 一阶最优性条件刻画了最优性 (设某一约束规范成立), 但 (从扰动分析的角度) 能刻画二阶增长条件的合适的推广仍然是很重要的.

这一节考虑问题

$$\text{(P)}\qquad \min_{x\in X} f(x)\quad \text{s. t.}\quad G(x)\in K, \tag{3.279}$$

设 X 与 Y 是 Banach 空间, f 与 G 是二次连续可微的, K 是 Y 的凸闭子集. 考虑可行集 $\Phi:=G^{-1}(K)$ 的闭子集 S, 满足对任何 $x\in S$, 有 $f(x)=f_0$ 是常值.

3.5.1 二次增长性的必要条件

下述定义给出相对于满足任何 $x\in S$ 均有 $f(x)=f_0$ 的闭集合$S\subset\Phi$ 的二阶增长条件的局部形式.

定义 3.141 *称局部二阶增长条件在点 $x_0\in S$ 处相对于集合 S 成立, 若存在 $\varepsilon>0, c>0$ 满足*

$$f(x)\geqslant f_0+c[\operatorname{dist}(x,S)]^2,\quad \forall x\in\Phi\cap B(x_0,\varepsilon). \tag{3.280}$$

当然, 若 x_0 是集合 S 的孤立点, 则上述条件 (3.280) 变为

$$f(x)\geqslant f(x_0)+c\|x-x_0\|^2,\quad \forall x\in\Phi\cap B(x_0,\varepsilon),$$

其中 $\varepsilon > 0$ 充分小, 此种情况, (3.280) 意味着二阶增长条件于 x_0 点成立. 不难验证, 若集合 S 是紧致的, 局部的二阶增长条件在 S 的每一点均成立, 则在定义 3.1 的意义下二阶增长条件在集合 S 处成立.

基于下述迫近法向量(proximal normal) 的概念, 讨论局部二阶增长性的二阶(必要的或充分的) 最优性条件.

定义 3.142 设 E 是 X 的闭子集, $x_0 \in E$.

(i) 称 $h \in X$ 是 E 在 x_0 处的迫近法向量, 若存在数 $\tau > 0$ 满足 $\operatorname{dist}(x_0+\tau h, E) = \tau\|h\|$. 用 $PN_E(x_0)$ 记 E 在 x_0 处的迫近法向量的集合.

(ii) 对某一常数 $\delta \geqslant 0$, 称 $h \in X$ 是 E 在 x_0 处的 δ 迫近法向量(δ-proximal normal), 若 $\operatorname{dist}(h, PN_E(x_0)) \leqslant \delta\|h\|$. E 在 x_0 处的 δ 迫近法向量的集合记为 $PN_E^\delta(x_0)$.

不难验证, 若 $h \in X$ 是 E 在 x_0 处的迫近法向量, 其中 $\tau > 0$ 是相应的常数, 则对任意 $t \in [0,\tau]$ 有 $\operatorname{dist}(x_0 + \tau h, E) = t\|h\|$. 由上面的定义有集合 $PN_E(x_0)$ 与 $PN_E^\delta(x_0)$ 均是锥, 可能只包含元素 $0 \in X$. 对 $\delta = 0$, 集合 $PN_E^\delta(x_0)$ 与 $PN_E(x_0)$ 重合. 注意到 $0 \in PN_E(x_0)$, 有 $\operatorname{dist}(h, PN_E(x_0)) \leqslant \|h\|$, 从而对 $\delta \geqslant 1$, 集合 $PN_E^\delta(x_0)$ 与整个空间 X 是重合的. 因此只考虑 $\delta \in [0,1)$ 时的 δ 迫近法向量才是有意义的.

由 2.2.4 节, E 在 $x \in E$ 处的余切 (外切) 锥定义为 $T_E(x) := \limsup_{t\downarrow 0} t^{-1}(E-x)$, 法锥$N_E(x_0) := [T_E(x_0)]^-$. 不难验证, 若 X 是 Hilbert 空间(有 X^* 与 X 是等同的), 则 $PN_E(x_0) \subset N_E(x_0)$, 若还有 E 是凸集, 则 $PN_E(x_0) = N_E(x_0)$.

例 3.143 考虑集合 $E := \{(x_1,x_2) \in \mathbb{R}^2 : x_2 \leqslant x_1^{4/3}\}, x_0 := (0,0)$. 得到 $T_E(x_0) = \{(x_1,x_2) : x_2 \leqslant 0\}$, 锥 $N_E(x_0)$ 由向量 $h := (0,1)$ 生成. 另一方面, $PN_E(x_0) = \{(0,0)\}$, 因此包含关系 $PN_E(x_0) \subset N_E(x_0)$ 是严格的.

下一个引理给出沿 δ 迫近法向量到 S 的距离的估计.

引理 3.144 对 $\delta \geqslant 0$, 令 h 是 S 在 x_0 处的 δ 迫近法向量, 则对充分小的 $t > 0$,

$$\operatorname{dist}(x_0 + th, S) \geqslant t(1-2\delta)\|h\|. \tag{3.281}$$

证明 令 $\varepsilon > 0$ 及 $h' \in PN_S(x_0)$ 满足

$$\|h - h'\| \leqslant (1+\varepsilon)\operatorname{dist}(h, PN_S(x_0)).$$

由于距离函数 $\operatorname{dist}(\cdot, S)$ 是模为 1 的 Lipschitz 连续, $\operatorname{dist}(h, PN_S(x_0)) \leqslant \delta\|h\|$, 对充分小的 $t \geqslant 0$, 有

$$\begin{aligned}\operatorname{dist}(x+th,S) &\geqslant \operatorname{dist}(x_0+th',S) - t\|h-h'\| \\ &= t(\|h'\| - \|h-h'\|) \geqslant t(\|h\| - 2\|h-h'\|) \\ &\geqslant t(1-2(1+\varepsilon)\delta)\|h\|.\end{aligned}$$

由于 ε 可取任意的小, 结论得证. □

定理 3.145 (二阶增长性的必要条件)　设在点 $x_0 \in S$ 处相对于 S 的二阶增长条件成立, 其中 $c > 0$ 与 $\varepsilon > 0$ 为相应的常数, 且 Robinson 约束规范在 x_0 处成立. 则对任何 $\delta \in \left(0, \frac{1}{2}\right)$, 存在 $\varepsilon' \in (0, \varepsilon)$ 满足对 $\forall x \in S \cap B(x_0, \varepsilon'), h \in C(x) \cap PN_S^\delta(x)$, 及任何凸集 $\mathcal{T}(h) \subset T_K^2(G(x), DG(x)h)$, 下述不等式成立

$$\sup_{\lambda \in \Lambda(x)} \left\{D_{xx}^2 L(x, \lambda)(h, h) - \sigma(\lambda, \mathcal{T}(h))\right\} \geqslant 2c(1 - 2\delta)^2 \|h\|^2. \tag{3.282}$$

证明　由于 Robinson 约束规范在 x_0 处成立, 则存在 $\varepsilon' \in (0, \varepsilon)$, 对所有的 $x \in S \cap B(x_0, \varepsilon')$ 这一规范也成立. 令 $x \in S \cap B(x_0, \varepsilon')$ 及 $h \in C(x) \cap PN_S^\delta(x)$, 设 $w \in X$ 是下述问题的可行点

$$\begin{aligned} &\min_{w \in X} Df(x)w + D^2 f(x)(h, h), \\ &\text{s.t.} DG(x)w + D^2 G(x)(h, h) \in \mathcal{T}(h). \end{aligned} \tag{3.283}$$

则存在形式为 $x_t := x + th + \frac{1}{2}t^2 w + o(t)$ 的一可行路径. 用局部二阶增长条件与引理 3.144, 对充分小的 $t \geqslant 0$, 有

$$f(x_t) - f(x) \geqslant c[\text{dist}(x_t, S)]^2 \geqslant ct^2(1 - 2\delta)^2 \|h\|^2 + o(t^2).$$

另一方面

$$f(x_t) - f(x) = \frac{1}{2}t^2 [Df(x)w + D^2 f(x)(h, h)] + o(t^2), \tag{3.284}$$

得到

$$Df(x)w + D^2 f(x)(h, h) \geqslant 2c(1 - 2\delta)^2 \|h\|^2,$$

对 w 取极小化得到问题 (3.283) 的最优值大于或等于 $2c(1 - 2\delta)^2 \|h\|^2$. 另一方面, 由于 Robinson 约束规范在 x 处成立, 由对偶性可得, 问题 (3.283) 的最优值等于 (3.282) 式的左端. 这证得 (3.282).　□

注 3.146　若 x_0 是 S 的孤立点, 则 $PN_S(x_0) = PN_S^\delta(x_0) = X$. 因此, 上述定理简化为定理 3.45.

在后面, 有时需要不必是 δ 追近法向量(对充分小的 δ) 的临界方向. 我们要构建对所有的临界方向均有效的某些必要性条件. 然而, 为此目的, 需要限定讨论在空间 X 是有限维的情况. 在有限维空间中, 余切锥 $T_S(x_0)$ 有下述性质.

引理 3.147　设空间 X 是有限维的, 令 $x_0 \in S, h \in X$. 则对 $t \geqslant 0$ 有

$$\text{dist}(x_0 + th, S) \geqslant t\ \text{dist}(h, T_S(x_0)) + o(t). \tag{3.285}$$

证明　给定 $\varepsilon > 0$①, 对 $t \geqslant 0$, 令 $x_t \in S$ 满足

$$\|x_0 + th - x_t\| \leqslant \text{dist}(x_0 + th, S) + o(t).$$

① 实际上原著中的证明没有用到 ε.

则 $t^{-1}(x_t - x_0)$ 是有界的, 这是因为

$$\begin{aligned}\|x_0 - x_t\| &\leqslant \|x_0 + th - x_t\| + t\|h\| \\ &\leqslant \operatorname{dist}(x_0 + th, S) + t\|h\| + o(t) \leqslant 2t\|h\| + o(t).\end{aligned}$$

令 $t_n \downarrow 0$ 是一序列, 沿此序列 $t^{-1}\operatorname{dist}(x_0 + th, S)$ 取到下极限. 如有必要, 可取子序列, 设 $t_n^{-1}(x_{t_n} - x_0)$ 具有极限 $\bar{h} \in T_S(x_0)$. 则

$$\begin{aligned}\operatorname{dist}(h, T_S(x_0)) \leqslant \|h - \bar{h}\| &= \lim_{n\to\infty} t_n^{-1}\|x_0 + t_n h - x_{t_n}\| \\ &= \liminf_{t\downarrow 0} t^{-1}\operatorname{dist}(x_0 + th, S),\end{aligned}$$

因此结论成立. □

定理 3.148 (二阶增长性的必要条件) *设空间 X 是有限维的, 局部二阶增长条件在点 $x_0 \in S$ 相对于 S 成立, 相应的常数 $c > 0, \varepsilon > 0$, Robinson 约束规范在 x_0 处成立. 则存在 $\varepsilon' \in (0, \varepsilon)$ 满足对所有的 $x \in S \cap B(x_0, \varepsilon'), h \in C(x)$ 及任何凸集合 $\mathcal{T}(h) \subset T_K^2(G(x), DG(x)h)$, 下述不等式成立*

$$\sup_{\lambda \in \Lambda(x)} \left\{D_{xx}^2 L(x, \lambda)(h, h) - \sigma(\lambda, \mathcal{T}(h))\right\} \geqslant 2c[\operatorname{dist}(h, T_S(x_0))]^2. \tag{3.286}$$

证明 由于在 x_0 处 Robinson 约束规范成立, 存在 $\varepsilon' \in (0, \varepsilon)$, 该规范对所有的 $x \in S \cap B(x_0, \varepsilon')$ 处均是成立的.

令 $w \in X$ 是问题 (3.283) 中的可行点. 则存在形式为 $x_t = x + th + \dfrac{t^2}{2}w + o(t)$ 的可行路径. 用局部二阶增长条件与引理 3.144, 得到

$$f(x_t) - f(x_0) \geqslant c[\operatorname{dist}(x_t, S)]^2 \geqslant t^2 c[\operatorname{dist}(h, T_S(x_0))]^2 + o(t^2).$$

而另一方面, 由 (3.284) 得

$$Df(x)w + D^2 f(x)(h, h) \geqslant 2c[\operatorname{dist}(h, T_S(x_0))]^2.$$

对所有的可行点 w 取极小化得到, 问题 (3.283) 的最优值大于或等于上述不等式的右端. 由于 Robinson 约束规范在 x 处成立, 由对偶性理论得, (3.283) 的最优值等于 (3.286) 中的左端, 证得 (3.286). □

3.5.2 充分条件

现在转到二阶充分条件的研究. 点 $\hat{x} \in S$ 称为点 $x \in X$ 到 S 上的 (度量) 投影, 记为 $\hat{x} = P_S(x)$, 若 $\|x - \hat{x}\| = \operatorname{dist}(x, S)$. 若对 $\varepsilon > 0$ 充分小, 集合 $\{x \in S : \|x - x_0\| \leqslant \varepsilon\}$ 是紧致的 (尤其若 X 是有限维的), 这样的投影在 $x_0 \in S$ 的一邻域内

是存在的, 尽管不是唯一的. 由定理 2.31, 若 S 是凸集, X 是自反的Banach 空间, 则投影 $P_S(x)$ 亦是存在的.

定义 3.149　称在 $x_0 \in S$ 处相对于 S 的局部二阶充分条件成立, 若存在 $\eta > 0, \beta > 0$ 满足对充分接近于 x_0 的所有的 $x \in S$, 及所有的 $h \in C_\eta(x) \cap PN_S(x)$, 下述不等式成立

$$\sup_{(\alpha,\lambda)\in\Lambda_N^g(x)} D_{xx}^2 L^g(x,\alpha,\lambda)(h,h) \geqslant \beta\|h\|^2. \tag{3.287}$$

若 x_0 是 S 的孤立点, 则上述的局部二阶充分条件与二阶充分条件 (3.136) 是重合的. 此种情况, 下述定理简化为定理 3.63.

定理 3.150 (非孤立极小点的二阶充分性条件)　令 $x_0 \in S$ 满足 (i) 对充分接近于 x_0 的每一 $x \in \Phi, P_S(x)$ 存在, 且 (ii) 局部的二阶充分条件 (3.287) 在 x_0 处成立. 则在 x_0 处相对于 S 的局部的二阶增长性条件成立.

证明　该证明是定理 3.63 证明的推广. 若结论不成立, 则存在收敛到 x_0 的序列 $x_n \in \Phi\backslash S$ 满足

$$f(x_0) + o(t_n^2) \geqslant f(x_n). \tag{3.288}$$

令 $\hat{x}_n := P_S(x_0)$. 则 $\hat{x}_n \to x_0$, 若有必要, 可选取一子序列, 设 $x_n = \hat{x}_n + t_n h_n$ 其中 $\|h_n\| = 1, h_n \in PN_S(\hat{x}_n), t_n \downarrow 0$. 引理 3.59 证明的直接的推广表明, 对充分大的 n 有 $h_n \in C_\eta(\hat{x}_n)$. 由 (3.287) 得

$$D_{xx}^2 L^g(\hat{x}_n,\alpha_n,\lambda_n)(h_n,h_n) \geqslant \beta \tag{3.289}$$

对某一正规化的广义 Lagrange 乘子成立, 因此

$$L^g(x_n,\alpha_n,\lambda_n) - L^g(\hat{x}_n,\alpha_n,\lambda_n) = \frac{1}{2}t_n^2 D_{xx}^2 L(\hat{x}_n,\alpha_n,\lambda_n)(h_n,h_n) + o(t_n^2) \geqslant \frac{1}{2}t_n^2\beta,$$

其中, 由 (α_n,λ_n) 的有界性有 $o(t_n^2)/t_n^2 \to 0$. 另一方面, 由于 $\lambda_n \in N_K(G(\hat{x}_n))$ 且 $G(x_n) \in K$, 因此 $\langle\lambda_n, G(x_n) - G(\hat{x}_n)\rangle \leqslant 0$, 由 (3.288) 可推得

$$L^g(x_n,\alpha_n,\lambda_n) - L^g(\hat{x}_n,\alpha_n,\lambda_n) \leqslant \alpha_n(f(x_n) - f(\hat{x}_n)) \leqslant o(t_n^2),$$

这导致相应的矛盾. □

基于精确临界方向的二阶条件, 我们叙述上述结果的有用的变形.

定义 3.151　称在 $x_0 \in S$ 处的临界锥的一致近似性质成立, 若对每一 $\varepsilon > 0$, 存在 $\eta > 0$ 满足对所有充分接近于 x_0 的 $x \in S$ 及任何 $h \in C_\eta(x) \cap PN_S(x)$, 则存在 $\hat{h} \in C(x)$ 满足 $\|\hat{h} - h\| \leqslant \varepsilon\|h\|$.

由上述定义得 $\|\hat{h}\| \geqslant (1-\varepsilon)\|h\|$, 从而 $\|\hat{h} - h\| \leqslant \delta\|\hat{h}\|$, 其中 $\delta := \varepsilon(1-\varepsilon)^{-1}$, 因此 $\hat{h} \in PN_S^\delta(x)$.

推论 3.152 令 S 是 Φ 的子集, 其上 $f(x)$ 是常值, 令 $x_0 \in S$ 满足

(i) 对充分接近 x_0 的每一 $x \in \Phi, P_S(x)$ 存在.

(ii) $x_0 \in S$ 处的临界锥的一致近似性质成立.

(iii) 下述二阶充分条件成立: 存在常数 $\beta > 0, \delta \in (0,1)$, 对充分接近于 x_0 的所有的 $x \in S$ 及 $h \in C(x) \cap PN_S^{\delta}(x)$, 下述不等式成立

$$\sup_{(\alpha,\lambda)\in\Lambda_N^g(x)} D_{xx}^2 L^g(x,\alpha,\lambda)(h,h) \geqslant \beta\|h\|^2. \tag{3.290}$$

则 x_0 处的二阶增长条件成立.

证明 只需要注意到, 由在 $x_0 \in S$ 处的临界锥的一致近似性质, 不等式 (3.290) 可推出 (3.287), 应用定理 3.150 可得结论. □

注 3.153 令 $S = \{x_0\}$, 其中 x_0 是下述问题的一可行点

$$\min_{x\in X} f(x) \quad \text{s.t.} \ g_i(x) = 0, \quad i = 1, \cdots, q; \quad g_i(x) \leqslant 0, \quad i = q+1, \cdots, p,$$

其中 f 与 g_i, $i = 1, \cdots, p$ 是 C^2 光滑函数. 则由 Hoffman 引理(定理 2.200), x_0 处的临界锥的一致近似性质成立, 因此, 在 $x = x_0$ 处的条件 (3.290) 是二阶增长性的充分条件 (若约束规范成立, 它还是必要的).

为了缩小二阶必要条件与二阶充分条件间的差别, 需考虑集合 K 的曲率. 在约束规范及临界锥的一致近似性质以及下述的一致二阶正则性 (与二阶正则性的定义 3.85 相比较) 假设之下, 可以得到用二阶条件表述的局部二阶增长条件的刻画.

定义 3.154 称 K 在 $G(x_0)$ 附近关于映射 G 与集合 S 是一致二阶正则的, 若对 x_0 的邻域 N_{x_0}, 对 $x \in S \cap N_{x_0}$ 与 $h \in C(x)$, 二阶切集 $T_K^{i,2}(G(x), DG(x)h)$ 在 $S \cap N_{x_0}$ 上一致地为 K 在点 $G(x)$ 沿方向 $DG(x)h$ 相对于 $DG(x)$ 的上二阶近似. 即若 $x_k \in S \cap N_{x_0}, h_k \in C(x_k), t_k \downarrow 0, r_k = DG(x_k)z_k + a_k$ 是满足 $\{a_k\}$ 收敛, $t_k z_k \to 0$ 及 $G(x_k) + t_k DG(x_k)h_k + \frac{1}{2}t_k^2 r_k \in K$ 的序列, 则

$$\lim_{k\to\infty} \operatorname{dist}(r_k, T_K^{i,2}(G(x_k), DG(x_k)h_k)) = 0. \tag{3.291}$$

定理 3.155 设 X 是有限维空间, $x_0 \in S$. 设 (i)Robinson 约束规范在 x_0 处成立; (ii) x_0 处的临界锥的一致近似性质成立; (iii) 集合 K 在 $G(x_0)$ 附近关于映射 G 与集合 S 是一致二阶正则的. 则 x_0 处的局部二阶增长条件成立的充分必要条件是对某一 $\delta \in \left(0, \frac{1}{2}\right)$, 存在 $\beta > 0$, 满足对充分接近于 x_0 的所有 $x \in S$ 及 $h \in C(x) \cap PN_S^{\delta}(x)$, 下述不等式成立

$$\sup_{\lambda\in\Lambda(x)} \left\{ D_{xx}^2 L(x,\lambda)(h,h) - \sigma(\lambda, T_K^{i,2}(G(x), DG(x)h)) \right\} \geqslant \beta\|h\|^2. \tag{3.292}$$

证明 (3.292) 的必要性由定理 3.145 可得, 充分性可如下证明. 若局部二阶增长条件不成立, 则可以像定理 3.148 的证明那样构造序列 $x_n, \hat{x}_n$ 与 h_n 满足 (3.288). h_n 到 $C(\hat{x}_n)$ 的投影 $\hat{h}_n$ 满足 $\|\hat{h}_n - h_n\| \to 0$, 因此有

$$x_n = \hat{x}_n + t_n\hat{h}_n + \frac{1}{2}t_n^2 w_n,$$

满足 $t_n w_n \to 0$. 从而有

$$G(x_n) = G(\hat{x}_n) + t_n DG(\hat{x}_n)\hat{h}_n + \frac{1}{2}t_n^2(DG(\hat{x}_n)w_n + D^2G(x_0)(\hat{h}_n, \hat{h}_n)) + o(t_n^2).$$

将 f 同样地展开, 用二阶正则性, 对某一序列 $\varepsilon_n \downarrow 0$, 有

$$\begin{aligned}&Df(\hat{x}_n)w_n + D^2f(\hat{x}_n)(\hat{h}_n, \hat{h}_n) \leqslant \varepsilon_n,\\&DG(\hat{x}_n)w_n + D^2G(\hat{x}_n)(\hat{h}_n, \hat{h}_n) \in T_K^{i,2}(G(\hat{x}_n), DG(\hat{x}_n)\hat{h}_n) + \varepsilon_n B_Y,\end{aligned}$$

于是

$$\begin{aligned}0 \geqslant\ & Df(\hat{x}_n)w_n + D^2f(\hat{x}_n)(\hat{h}_n, \hat{h}_n) - \varepsilon_n\\&+\langle \lambda, DG(\hat{x}_n)w_n + D^2G(\hat{x}_n)(\hat{h}_n, \hat{h}_n)\rangle\\&-\sigma(\lambda, T_K^{i,2}(G(\hat{x}_n), DG(\hat{x}_n)\hat{h}_n)) - \varepsilon_n\|\lambda\|\\=\ & D_{xx}^2(\hat{x}_n, \lambda)(\hat{h}_n, \hat{h}_n) - \sigma(\lambda, T_K^{i,2}(G(\hat{x}_n), DG(\hat{x}_n)\hat{h}_n)) - \varepsilon_n(\alpha + \|\lambda\|),\end{aligned}$$

这与 (3.292) 相矛盾. □

注 3.156 设临界锥的一致近似性质成立, 则可用上述证明的方法, 不难利用"一致"上二阶近似与广义 Lagrange 乘子来描述二阶充分性条件.

问题 (P) 被称为有限约束的, 若其可行集 Φ 由有限多个等式与不等式约束定义, 即

$$\Phi := \{x : g_i(x) = 0, i = 1, \cdots, q;\quad g_i(x) \leqslant 0, i = q+1, \cdots, p\}. \tag{3.293}$$

用 $I(x)$ 记在 x 处起作用的不等式约束的指标集, 即

$$I(x) := \{i : g_i(x) = 0,\ i = q+1, \cdots, p\},$$

也记 $I^*(x) := \{1, \cdots, q\} \cup I(x)$.

若问题 (P) 是有限约束的, 则局部二阶正则性条件是显然成立的, 因为此种情形 $T_K^{i,2}(y, d) = T_{T_K(y)}(d)$. 所以, 我们得到下述推论.

推论 3.157 *设 X 是有限维空间, $x_0 \in S$. 设*

(i) 问题 (P) 是有限约束的.

(ii) Mangasarian-Fromovitz 约束规范在 x_0 处成立.

(iii) 临界锥的一致近似性质在 x_0 处成立.

则 x_0 处的局部二阶增长条件成立的充分必要条件是对某 $\delta \in \left(0, \frac{1}{2}\right)$, 存在 $\beta > 0$, 满足对充分接近于 x_0 的所有 $x, h \in C(x) \cap PN_S^\delta(x)$, 下述不等式成立

$$\sup_{\lambda \in \Lambda(x)} D_{xx}^2 L(x, \lambda)(h, h) \geqslant \beta \|h\|^2. \tag{3.294}$$

下述例子表明, 对非线性规划问题而言, 临界锥的一致近似性质的条件不一定成立.

例 3.158 考虑优化问题

$$\min_{x \in \mathbb{R}^2} \ x_1 x_2 \quad \text{s.t.} \ x_1 \geqslant 0, \quad x_2 \geqslant 0. \tag{3.295}$$

上述问题的最优解集是

$$S = (\mathbb{R}_+ \times \{0\}) \cup (\{0\} \times \mathbb{R}_+).$$

对任何 $x \in \mathbb{R}_+ \times \{0\}, x \neq 0$, 临界锥 $C(x)$ 等于 $\mathbb{R} \times \{0\}$, 对 $x \in \{0\} \times \mathbb{R}_+$ 也有类似的结果. 然而, 对任何给定的 $\eta > 0$, 若 $x \in \mathbb{R}_+ \times \{0\}, x \neq 0$ 充分地接近于 0, 则有 $C_\eta(x) = \mathbb{R} \times \mathbb{R}_+, PN_S(x) = \{0\} \times \mathbb{R}_+$, 因此 $C_\eta(x) \cap PN_S(x) = \{0\} \times \mathbb{R}_+$. 结果, 临界锥的一致近似的条件是不成立的.

尽管上述定理是孤立最优解的相应结果的自然推广, 然而这些结果是不容易应用的, 因为我们对确保临界锥的一致近似性与局部二阶正则性成立的充分条件几乎不知道什么. 尽管这样, 对非线性规划的情况, 在两个特别的情形下, 仍然可叙述临界锥的一致近似性的某些充分条件.

命题 3.159 设 X 是有限维空间, (P) 是有限约束问题, 即其可行集是如 (3.293) 中由有限多个约束定义. 设梯度 $\nabla g_i(x_0), i \in I^*(x_0)$ 是线性无关的, 严格互补条件在 x_0 成立, 即与 x_0 相联系的 (唯一的)Lagrange 乘子 $\lambda(x_0)$ 满足对所有的 $i \in I(x)$①, 有 $\lambda_i(x_0) > 0$. 则临界锥的一致近似性质是成立的.

证明 对与 x_0 充分接近的 $x \in S$, 有 $I(x) \subset I(x_0)$. 由于 $\nabla g_i(x_0), i \in I^*(x_0)$ 是线性无关的, 这意味着与 x 相联系着唯一的 $\lambda(x)$, 且当 $x \to x_0$ 时有 $\lambda(x) \to \lambda(x_0)$. 由严格互补性质, 对充分接近于 x_0 的 x, $I(x) = I(x_0)$ 且在 x 处严格互补性质亦是成立的, 因此有

$$C(x) = \{h \in X : Dg_i(x)h = 0, i \in I^*(x_0)\}.$$

因为分析是局部的, 可以设所有的约束在 x_0 处均是起作用的. 置 $A(x) := DG(x)$. 则 h 到 $C(x)$ 上的直交投影 $P(x)$ 为

① 应该是 $I(x_0)$.

$$P(x)h = h - A(x)^{\mathrm{T}}(A(x)A(x)^{\mathrm{T}})^{-1}A(x)h.$$

所以, 在 x_0 的邻域内, 有

$$\|P(x)h - h\| = \|A(x)^{\mathrm{T}}(A(x)A(x)^{\mathrm{T}})^{-1}A(x)h\| \leqslant \gamma\|A(x)h\|, \tag{3.296}$$

其中 $\gamma := 2\|A(x_0)^{\mathrm{T}}(A(x_0)A(x_0)^{\mathrm{T}})^{-1}\|$. 为了记号方便, 设问题没有等式约束, 只需证明, 存在 $c > 0$, 对充分接近于 x_0 的 x, 下述不等式成立

$$\|A(x)h\| \leqslant c([Df(x)h]_+ + [DG(x)h]_+). \tag{3.297}$$

假设这是不真的, 则存在序列 $x_k \to x_0$ 与 h_k 满足

$$\|A(x_k)h_k\| > k([Df(x_k)h_k]_+ + [DG(x_k)h_k]_+). \tag{3.298}$$

若有必要, 将 h_k 替换为 $h_k/\|A(x_k)h_k\|$, 可设 $\|A(x_k)h_k\|$ 等于 1. 因为 $A(x)$ 是 x 的连续函数, 它在 x_0 的邻域内是满秩的, 容易构造有界序列 $\{\hat{h}_k\}$ 满足 $A(x_k)h_k = A(x_k)\hat{h}_k$, 因此, 由一阶最优性系统, $Df(x_k)$ 是 $DG(x_k)$ 的列的线性组合, $Df(x_k)h_k = Df(x_k)\hat{h}_k$.

设 $\{\hat{h}_k\}$ 具有极限点h_0. 由 (3.298) 得 $\|A(x_0)h_0\| = 1, Df(x_0)h_0 \leqslant 0$ 且 $DG(x_0)h_0 \leqslant 0$, 因此 h_0 是满足 $A(x_0)h_0 = DG(x_0)h_0 \neq 0$ 的非零的临界方向. 由于对所有的临界方向 h, $\langle\lambda(x_0), DG(x_0)h\rangle = 0, DG(x_0)h_0 \leqslant 0, \lambda(x_0)_i > 0, \forall i$, 这导致矛盾. □

命题 3.160　设 X 是有限维的空间, (P) 是有限约束的凸问题, 即其可行集如 (3.293) 所示由有限多个约束来定义, 其中函数 f 与 $g_i, i = q+1, \cdots, p$ 是凸的连续的, $g_i, i = 1, \cdots, q$ 是仿射的, 设最优解集 $S := \mathcal{S}(P)$ 是非空的. 则临界锥的局部一致近似性质是成立的.

证明　不失一般性, 设 $q = 0$ (即没有等式约束存在). 证明分几步.

(a) 因为 (P) 是凸的, 集合 S 也是凸的. 我们给出论断: 存在 I, C, 与 T 满足对所有的 $x \in \mathrm{ri}(S)$, 有 $I(x) = I, C(x) = C, T_S(x) = T$. 事实上, 首先证明, 当 $x \in \mathrm{ri}(S)$ 时, 则 $I(x) = \bigcap\limits_{x' \in S} I(x')$, 因而是常值的. 假设 $i \in I(x)$, 对某一 $x' \in S, i \notin I(x')$. 则由于 g_i 是凸的,

$$Dg_i(x)(x' - x) \leqslant g_i(x') - g_i(x) < 0.$$

得到 $Dg_i(x)(x - x') > 0$. 所以, 对充分小的 $\varepsilon > 0$, 有 $g_i(x + \varepsilon(x - x')) > 0$, 这同 $x + \varepsilon(x - x') \in S$ 矛盾. 我们证明 $C(x)$ 在 $\mathrm{ri}(S)$ 是常值的. 考虑某一 $i \in I$. 由命题 2.119 可得 $\partial g_i(x)$ 在 $\mathrm{ri}(S)$ 上是常值的. 结果, $Dg_i(x)$ 在 $\mathrm{ri}(S)$ 上是常值的. 同样的结论应用于 f. 就得到 $C(x)$ 在 $\mathrm{ri}(S)$ 上是常值的. 对于 $T_S(x)$, 结论是平凡的, 因为在每一 $x \in \mathrm{ri}(S)$ 处它是平行于 $\mathrm{Sp}(S)$ 的线性子空间.

(b) 因为在 S 的相对内部上临界锥是常值的, 若 $x_0 \in \mathrm{ri}(S)$, 则结果由 Hoffman 引理(定理 2.200) 可得到.

(c) 更进一步考虑集合 C. 置 $g_0(x) := f(x)$, 可表示

$$C = \{h \in X : A^{\mathrm{T}} h \leqslant 0\},$$

其中 A 是矩阵, 它的列 a_i 由 $\nabla g_i(x), i \in \{0, \cdots, q\} \cup I$ 给出, 对任何 $x \in \mathrm{ri}(S)$. 置

$$I_0 := \{i \in I : a_i h = 0, \forall h \in C\}, \quad I_1 := I \backslash I_0.$$

同样定义 $A_0 := (a_i)_{i \in I_0}, A_1 := (a_i)_{i \in I_1}$, 从而 $A = A_0 \cup A_1$. 给出论断, 存在 $h_0 \in C$ 满足 $A_1 h_0 < 0$. 实际上, 对每一 $i \in I_1$, 联系着 $h_i \in C$ 满足 $a_i h_i < 0$, 从而 $h_0 := \sum_{i \in I_1} h_i$ 是这样的向量.

(d) 分析当 x 属于集合 S 的相对边界时的 $C(x)$. 定义

$$J(x) := I(x) \backslash I, \quad B(x) := (Dg_i(x))_{i \in J(x)}.$$

令 $x' \in \mathrm{ri}(S), i \in J(x)$. 则 $Dg_i(x)(x' - x) \leqslant g_i(x') - g_i(x) < 0$. 方向 $h := x' - x$ 属于 $T_S(x)$, 因此属于 $C(x)$. 因此, 若 $J(x) \neq \varnothing$, 则

$$C(x) = \{h : Ad \leqslant 0, B(x)d \leqslant 0\},$$

且存在 $h_1 \in C(x)$ 满足 $B(x)h_1 < 0$.

由步 (c) 及包含关系 $C(x) \subset C$, 存在 $\varepsilon > 0$ 满足 $h_2 := h_1 + \varepsilon h_0$ 满足 $A_1 h_2 < 0$ 与 $B(x)h_2 < 0$. 对任何 $J \subset J(x)$, 置 $B_J(x) := (Dg_i(x))_{i \in J}$ 与

$$H_J(x) := \{h : A_0 h = 0, A_1 h \leqslant 0, B_J(x)h \leqslant 0\}.$$

知 $h_2 \in H_J(x)$ 满足 $A_1 h_2 < 0$ 且 $B_J(x)h_2 < 0$. 可设 A_0 是一组无关向量的集合, 因为去掉线性相关的向量不会改变它的核. 观察到 $B_J(x)$ 连续地依赖于 x. 由稳定性定理 (定理 2.87), 存在 x 的邻域 U 及 $\delta > 0$ 满足

$$\mathrm{dist}(d, H_J(x')) \leqslant \delta(\|A_0 h\| + \|(A_1 d)_+\| + \|(B_J(x')d)_+\|) \tag{3.299}$$

对所有的 $h \in \mathbb{R}^n$ 与 x 的邻域 $\mathcal{V}$ 的所有的 x' 均成立.

进一步观察到, 存在 $\gamma' > 0$ 满足 $\|A_0 d\| \leqslant \gamma' \, \mathrm{dist}(d, \mathrm{ker} A_0)$, (3.299) 连同明显的包含关系 $C \subset \mathrm{ker} A_0$, 推出对某一 $\gamma = \gamma\ (x, J)$, 有

$$\mathrm{dist}(d, H_J(x')) \leqslant \gamma\ (\|(Ad)_+\| + \|(B_J(x')d)_+\|), \quad \forall d, \forall x' \in \mathcal{V}. \tag{3.300}$$

(e) 若结论不成立, 则可找到序列 $x_k \in S$ 及非零的向量序列 $h_k \in \mathbb{R}^n$ 满足

$$\operatorname{dist}(h_k, C(x_k)) \geqslant k \max_{i \in I_0(x_k)} (DG_i(x_k)(h_k))_+. \tag{3.301}$$

若有必要, 可抽取一子列, 可设 x_k 收敛到 $x_0 \in S, J(x_k)$ 等于某一 J. 由 (3.300), 存在 $\gamma > 0$ 满足对充分大的 k,

$$\operatorname{dist}(h_k, H_J(x_k)) \leqslant \gamma\, (\|(Ad)_+\| + \|(B_J(x_k)h_k)_+\|).$$

由于 $J = J(x_k)$, 有 $I \cup J = I(x_k), H_J(x_k) = C(x_k)$, 由 (3.300) 推出, 对某一 $\gamma_1 > 0$,

$$\operatorname{dist}(h_k, C(x_k)) \leqslant \gamma_1 \max_{i \in I_0(x_k)} (DG_i(x_k)(h_k))_+,$$

这与 (3.301) 是矛盾的. 这就完成了证明. □

处理非孤立最优解的另一方式是假定与 Lagrange 函数的一阶及二阶的展式相关联的某些条件成立. 对常数 $\varepsilon \geqslant 0$, 考虑如下定义的广义 Lagrange ε 乘子的集合 (与 3.1.3 节中的满足条件 (3.33) 的 "近似的"Lagrange 乘子相比较)

$$\Lambda_\varepsilon^g(x) := \{(\alpha, \lambda) \in \mathbb{R}_+ \times N_K(G(x)) : (\alpha, \lambda) \neq 0,\ \|D_x L^g(x, \alpha, \lambda)\| \leqslant \varepsilon\}. \tag{3.302}$$

对 $\varepsilon = 0$, 上述集合与广义 Lagrange 乘子的集合是相同的.

定义 3.161 *设 S 是 Φ 的子集, f 在 S 上取常值. 称 (对 S 的) 在 $x_0 \in S$ 处的广义的局部二阶充分条件成立, 若存在 $\eta > 0, \varepsilon > 0$ 与 $\beta > 0$, 满足对充分接近于 x_0 的所有 $x \in S$, $h \in C_\eta(x) \cap PN_S(x)$ 充分接近于零, 下述不等式成立*

$$\sup_{\substack{|\alpha| + \|\lambda\| \leqslant 1 \\ (\alpha, \lambda) \in \Lambda_\varepsilon^g(x)}} \left\{ D_x L^g(x, \alpha, \lambda)h + D_{xx}^2 L^g(x, \alpha, \lambda)(h, h) \right\} \geqslant \beta \|h\|^2. \tag{3.303}$$

若 x_0 是 S 的孤立点, 上述二阶条件简化为标准的二阶条件 (3.136).

定理 3.162 (基于 Lagrange 近似乘子的二阶充分性条件) *令 S 是 Φ 中的子集, f 在其上取常值. 令 $x_0 \in S$ 满足 (i) 每一充分接近于 x_0 的 $x \in \Phi$ 在 S 上均有投影; (ii) 在 $x_0 \in S$ 处广义局部二阶充分性条件 (3.303) 成立. 则 x_0 处的局部二阶增长条件成立.*

证明 其证明是定理 3.150 证明的一个很容易的推广. □

因为广义局部二阶充分条件 (3.287) 可推出局部的二阶充分条件 (3.303), 定理 3.162 是定理 3.150 的一个推广.

3.5.3 基于一般临界方向的充分性条件

例 3.158 表明, 对于非凸的非线性规划问题, 前一节定理的假设 (除了最后一定理, 它涉及更多的条件), 其条件经常不被满足. 这启发我们考虑最优解集中的 "接近" 点而不是 (度量)投影.

定理 3.163 设 S 是 Φ 的子集, f 于其上是常值的, $x_0 \in S$. 设存在正常数 ε, c 与 β①, 满足对充分接近于 x_0 的满足 $f(x) \leqslant f(x_0) + \varepsilon\, \mathrm{dist}(x, S)$ 的任何 $x \in \Phi$, 存在 $\hat{x} \in S$ 满足 $\|\hat{x} - x\| \leqslant \beta\ \mathrm{dist}(x, S)$ 且 $h := x - \hat{x}$ 满足

$$\sup_{(\alpha,\lambda)\in\Lambda_N^g(x)} D_{xx}^2 L^g(\hat{x}, \alpha, \lambda)(h, h) \geqslant 2c'[\mathrm{dist}(x, S)]^2. \tag{3.304}$$

则对任何常数 $c < c'$, x_0 处的局部的二阶增长条件成立.

证明 其证明是定理 3.150 证明的一个很容易的推广. □

上述定理的假设似乎太复杂和人工化了. 然而可以给出一些与可行集及最优解集相联系的几何条件, 在这些条件下, 若 "象征性的" 点 x_0 满足约束规范, 这些充分条件给出二阶增长性的刻画. 进一步, 下面的分析给出 $\hat{x}$ 的显式的简单构造.

定义 3.164 设 S_1 与 S_2 是 X 的子集, $x_0 \in S_1 \cap S_2$. 称 S_1 与 S_2 在 x_0 处没有 (非零的) 公共切向量, 若 $T_{S_1}(x_0) \cap T_{S_2}(x_0) = \{0\}$.

称集合与凸集是同构的若它是一个凸集在某一 C^2 类微分同胚下的像.

定理 3.165 设空间 X 是有限维的, (P) 是有限约束的, $x_0 \in S$ 满足

(i) Robinson 约束规范在 x_0 处成立.

(ii) 在 x_0 的邻域, 集合 S 可以表示为 $S_i, i = 1, \cdots, q$ 的并, 每一集合 S_i 与一凸集同构, 满足若 $1 \leqslant i < j \leqslant q$, 则 S_i 与 S_j 没有公共的切向量.

(iii) 若 $\hat{x}_k \to x_0, \{x_k\}$②包含在某一 S_i 中, $j \in I(x_0)\backslash I(\hat{x}_k)$, 则 $Dg_j(x_0) \neq 0$, 且 $g_j^{-1}(0)$ 与 S_i 没有公共的切向量.

则定理 3.163 的条件是 x_0 处的二阶增长条件的充分必要条件.

证明 由定理 3.163, 条件是充分的, 剩下要证这些条件还是必要的.

如下述固定 β 值. 对所有 $\varepsilon > 0$, 定义 S_l 在 x_0 处的 ε 切方向集合

$$T_{S_l}^{\varepsilon}(x_0) := \{h \in X : \mathrm{dist}(h, T_{S_l}(x_0)) \leqslant \varepsilon\|h\|\}. \tag{3.305}$$

对充分小的 $\varepsilon > 0$, 令 $h \in T_{S_l}^{\varepsilon}(x_0)$, $\hat{h}$ 是 h 到凸集 $T_{S_l}(x_0)$ 上的投影. 则由于 S③与凸集合同构, 有

$$\mathrm{dist}(x_0 + th, S_l) = t\|h - \hat{h}\| + o(t) \leqslant t\varepsilon\|h\| + o(t). \tag{3.306}$$

① 此处的 c 应该为 c'.

② 应该是 $\{\hat{x}_k\}$.

③ 应该是 S_l.

把这一关系式与引理 3.147 结合, 对充分小的 $t>0$, 可得到

$$\operatorname{dist}(x_0+th,S)=\operatorname{dist}(x_0+th,S_l). \tag{3.307}$$

设 $i\in I(x_0), x\in S_l$, 且 $i\notin I(x)$. 由 (iii), 设 x 充分接近于 x_0, 如需要可减少 ε, 有 $Dg_i(x_0)(x-x_0)\neq 0$, 且由 $g_i(x)\leqslant 0$ 得 $Dg_i(x_0)(x-x_0)<0$. 可取 $\varepsilon>0$ 充分小使得对任何可能的选择 i 与 l, 这一结果是真的.

令 $x\in X$ 到 S 的投影, 即 $P_S(x)$, 落于 S_l, 其中 l 是某指标. 置 $t:=\operatorname{dist}(x,S)=\|x-P_S(x)\|, \tau:=\|x_0-P_S(x)\|$.

给出论断：存在 $\beta>0$ 满足：若 $\beta t\leqslant\tau$, 则点

$$\hat{x}:=(1-\beta t/\tau)P_S(x)+(\beta t/\tau)x_0 \tag{3.308}$$

(它属于 $[x_0,P_S(x)]$) 满足 $h:=x-\hat{x}$ 属于 $T^{\varepsilon}_{S_l}(x_0)$. 实际上, 由 Pythagorean 定理,

$$\begin{aligned}\|h\|&=(\|x-P_S(x)\|^2+\|P_S(x)-\hat{x}\|^2)^{1/2}\\&\geqslant\|P_S(x)-\hat{x}\ \|=\frac{\beta t}{\tau}\|P_S(x)-x_0\|=\beta t,\end{aligned}$$

因而

$$\operatorname{dist}(h,T_{S_l}(x_0))\leqslant\|h-(P_S(x)-\hat{x})\|=\|x-P_S(x)\|=t\leqslant\beta^{-1}\|h\|,$$

于是可取 $\beta:=\varepsilon^{-1}$.

现在考虑满足 $f(x_k)\leqslant f(x_0)+o(\operatorname{dist}(x_0,x_k))$ 的序列 $\{x_k\}$. 用 $P_S(x_k)$ 记 x_k 到 S 的直交投影 (由关于 S 的假设, 对充分大的 k, 这一投影是存在的). 如有必要可抽取子列, 设 $\{P_S(x_k)\}$ 包含于某一 S_l 中. 置 $t_k:=\operatorname{dist}(x_k,S_l)$ 与 $\tau_k:=\|x_k-x_0\|$. 先设 S_l 是凸的.

情况 (a). 若 $t_k>\beta^{-1}\tau_k$, 则置 $\hat{x}:=x_0, h_k:=t_k^{-1}(x_k-x_0)$. 若有必要可抽取一子列, 设 h_k 具有极限 h_0. 由于 $\|h_k\|\geqslant 1$, 必有 $h_0\neq 0$. 由

$$0\geqslant g_i(x_k)=g_i(x_0)+t_kDg_i(x_0)h_k+o(t_k)=g_i(x_0)+t_kDg_i(x_0)h_0+o(t_k)$$

可推得, 对所有的 $i\in I(x_0)$ 有 $Dg_i(x_0)h_0\leqslant 0$. 由关系式 $f(x_k)\leqslant f(x_0)+o(\|x_k-x_0\|)$, 有 $Df(x_0)h_0\leqslant 0$, 从而 $h_0\in C(x_0)$. 由于 $x_k-P_S(x_k)$ 是 S_l 的朝外的法向量, 对 $\rho>0$ 有

$$\begin{aligned}\operatorname{dist}(x_0+\rho h_0,S_l)&=\operatorname{dist}(x_0+\rho h_k,S_l)+o(\rho)\\&=\operatorname{dist}(P_S(x_k)+\rho h_k,S_l)+o(\rho)=\rho+o(\rho).\end{aligned}$$

另一方面, 对充分小的 $\rho > 0, \operatorname{dist}(x_0 + \rho h_0, S) = \operatorname{dist}(x_0 + \rho h_0, S_l) + o(\rho)$, 因为否则就与 $\operatorname{dist}(x_k, S) = \operatorname{dist}(x_k, S_l)$ 这一事实相矛盾. 由定理 3.148, 条件 (3.287) 成立, 其中 c 即出现在二阶增长条件中的常数.

情况 (b). 若有必要可抽取一子列, 可假设 $t_k \leqslant \beta^{-1}\tau_k$. 考虑序列

$$\hat{x}_k := (1 - \beta t_k/\tau_k)P_S(x_k) + (\beta t_k/\tau_k)x_0.$$

由证明刚开始的讨论有 $h_k \in T_{S_l}^{\varepsilon}(x_0), Dg_i(x_0)h_k \leqslant 0, \forall i \in I(x_0)\backslash I(P_S(\hat{x}_k))$. 还有 $\hat{x}_k \in [x_0, P_S(x_k)]$ 且

$$\|x_k - \hat{x}_k\| = \|x_k - P_S(x_k) + (\beta t_k/\tau_k)(P_S(x_k) - x_0)\| \leqslant (1 + \beta)t_k.$$

于是 $h_k := t_k^{-1}(x_k - \hat{x}_k)$ 满足 $\|h_k\| \leqslant 1 + \beta$, 因而是有界的, 由于 $\|h_k\| \geqslant t_k^{-1}\|x_k - P_S(x_k)\| \geqslant 1$, 可设它收敛于某一 $h_0 \in T_{S_l}^{\varepsilon}(x_0)$ 满足 $1 \leqslant \|h_0\| \leqslant 1 + \beta$. 由于

$$f(x_k) \leqslant f(x_0) + o(\operatorname{dist}(x_k, S)) = f(\hat{x}_k) + o(\|x_k - \hat{x}_k\|),$$

有 $Df(x_0)h_0 \leqslant 0$. 还可以设 $I = I(\hat{x}_k)$ 是常值的, 由 $g(x_k)$ 的一阶展式, 有 $Dg_i(x_0)h_0 \leqslant 0, \forall i \in I$. 因为 $h_0 \in T_{S_l}^{\varepsilon}(x_0)$, 由证明开始时的结论, $Dg_i(x_0)h_0 \leqslant 0, i \in I(x_0) \setminus I$, 因此 h_k 是临界方向. 由定理 3.148, 用 $\operatorname{dist}(x_0 + t_k h_k, S)^2 = t_k^2$ 及 $\|h_k\| \leqslant 1 + \beta$, 得到

$$\begin{aligned}\sup_{\lambda \in \Lambda(x_0)} D_{xx}^2 L(x_0, \lambda)(t_k h_0, t_k h_0) &\geqslant 2c \operatorname{dist}(x_0 + t_k h, S)^2 \\ &\geqslant 2c \operatorname{dist}(x_0 + t_k h_k, S)^2 + o(t_k^2) \geqslant \frac{2ct_k^2}{1+\beta}\|h_0\|^2 + o(t_k^2).\end{aligned}$$

置 $c' := 1/(1 + \beta)$, 得结论成立.

若 S_i 不是凸的, 可以用 C^2 微分同胚的性质, 由它的像是凸的, 把分析减化到凸集的情形. 因为这里用到的各种假设在这样的变量变换下具有不变性, 结论得证. □

第 4 章　稳定性与灵敏度分析

这一章研究如下形式的参数化最优化问题

$$(\mathrm{P}_u) \qquad \min_{x\in X} f(x,u) \quad \text{s. t.} \quad G(x,u)\in K, \tag{4.1}$$

该问题依赖于参数向量 $u \in U$. 除非特别说明, 这一章假设 X, Y 与 U 是 Banach 空间, K 是 Y 中的闭凸子集, $f : X \times U \to \mathbb{R}, G : X \times U \to Y$ 是连续的. 在某些情况下, 我们处理 $f(\cdot, u)$ 在抽象集合 $\Phi(u) \subset X$ 中的极小化问题. 在任何情况下, 均用 $\Phi(u)$ 记目前最优化问题的可行集. 尤其, 对上述问题 (P_u), 有

$$\Phi(u) := \{x \in X : G(x,u) \in K\}. \tag{4.2}$$

对参数空间 U 中的给定的 u_0, 视相应的问题 (P_{u_0}) 为非扰动问题, 研究在点 u_0 的邻域中的最优值函数

$$v(u) := \inf_{x\in\Phi(u)} f(x,u) \tag{4.3}$$

与相应的最优解集

$$\mathcal{S}(u) := \operatorname*{argmin}_{x\in\Phi(u)} f(x,u) \tag{4.4}$$

的连续性质与可微性质. 我们也考虑 (P_u) 的近似 (ε 最优) 解. $\bar{x}$ 称为 (P_u) 的 ε 最优解, 若 $\bar{x} \in \Phi(u)$, 即 $\bar{x}$ 是可行的, 且 $f(\bar{x}, u) \leqslant v(u) + \varepsilon$.

经常视非扰动问题 (P_{u_0}) 与前面几节考虑的问题 (P) 相同. 即用 $f(\cdot) = f(\cdot, u_0)$, $G(\cdot) = G(\cdot, u_0)$, 可把非扰动问题记为下述形式

$$(\mathrm{P}) \qquad \min_{x\in X} f(x) \qquad \text{s. t.} \qquad G(x) \in K. \tag{4.5}$$

许多结果处理参数化问题 (P_u) 的方向分析, 即对给定的方向 $d \in U$, 研究沿具有形式为 $u(t) := u_0 + td + \varepsilon(t), t \geqslant 0$ 的路径上的最优值与最优解的稳定性质与微分性质, 其中 $\varepsilon(t)$ 可是 $o(t)$ 阶的或 $O(t^2)$ 阶的. 这是由于这样的事实, 即一般而言, 最优值函数与最优解不是处处可微的, 通过方向分析可以得到某些好的估计. 同时, 具有关于方向一致的稳定性结果是有用途的. 当可能的时候, 我们叙述这样的结果.

4.1 最优值与最优解的稳定性

这一节讨论参数化问题 (P_u) 的最优值函数 $v(u)$ 与最优解集 $\mathcal{S}(u)$ 的连续性质. 由定义, 若可行集 $\Phi(u)$ 是空集, 则 $v(u) = +\infty$. 不难给出例子, 即使目标函数 $f(x,u)$ 是连续的且可行集是固定的 (与 u 无关), 函数 $v(u)$ 也是不连续的.

例 4.1 考虑 $f(x,u) := \mathrm{e}^{-x^2u^2}, \Phi(u) := \mathbb{R}, x, u \in \mathbb{R}$. 则对任何 $u \neq 0, v(u) = 0, v(0) = 1$. 很清楚, 最优值函数在 $u = 0$ 处是不连续的. 这一例子中的 $v(u)$ 的这种病态现象与可行集的无界性相关, 这使得对于 $u \neq 0$, 最优解可能逃离到无穷远.

应该注意到, 若可行集是固定的, 如 $\Phi(u) = \Phi, \forall u \in U$, 则最优值函数 $v(u)$ 是上半连续的, 或等价地, $-v(u)$ 是下半连续的. 事实上, 有

$$-v(u) = \sup\{-f(x,u) : x \in \Phi\},$$

因此, $-v(\cdot)$ 的上图由 $-f(x,\cdot), x \in \Phi$ 的上图的交给出. 因为 $f(x,\cdot)$ 是连续的, $-f(x,\cdot)$ 的上图是闭的. 从而 $-v(\cdot)$ 的上图是闭集, $-v(\cdot)$ 是下半连续的.

在后面的叙述中, 设 X 与 U 为 Hausdorff 拓扑空间. 可行集 $\Phi(u)$ 与最优解集 $\mathcal{S}(u)$ 可视为从 U 到 2^X 的多值函数. 多值函数 $F : U \to 2^X$ 被称为是闭的, 若它的图 $\mathrm{gph}(F)$ 是 $U \times X$ 中的闭子集 (见 2.3 节). 多值函数 F 被称为是闭值的, 若对每一 $u \in U, F(u)$ 是 X 中的闭子集. 多值函数 F 称为在 $u_0 \in U$ 处上半连续的, 若对集合 $F(u_0)$ 的任何邻域 V_X, 存在 u_0 的邻域 V_U 满足对每一 $u \in V_U$, 包含关系 $F(U) \subset V_X$ 成立. 若这一性质在每一 $u_0 \in U$ 均是成立的, 则称 F 是上半连续的.

集合 $V \subset X$ 称为 X 中的子集 S 的邻域, 若 $S \subset \mathrm{int}(V)$. 在下面的分析中, 需要假设 X 的拓扑具有下述性质. 称一点的邻域包含闭邻域, 若 X 中的任意点的任何邻域均包含这一点的一个闭邻域. 如下述引理所示, 对本书中考虑的任何有意思的拓扑, 这一性质均是成立的.

引理 4.2 (i) *令 X 是紧致的 Hausdorff 拓扑空间, 或是度量空间, 或是局部凸的拓扑向量空间. 则 X 的拓扑具有性质: 一点的邻域包含一闭的邻域.*

(ii) *令 X 是满足一点的邻域包含闭的邻域的 Hausdorff 拓扑空间, 令 S 是 X 中的闭子集, $x \in X \setminus S$. 则存在 x 与 S 的邻域 V 与 W 满足 $V \cap W = \varnothing$.*

证明 (i) 设 X 是紧致的 Hausdorff 拓扑空间, V 是 x 的邻域. 则存在开集 $W \subset X$ 满足 $x \in W \subset V$. 由于 $X \setminus W$ 是闭的, 它是紧致集 X 中的子集, 因而亦是紧致的. 由分离公理, 对 $X \setminus W$ 中的任何 y, 存在不相交的开集 V_y 与 V'_y 分别包含 x 与 y. 由于是紧致的, $X \setminus W$ 被 $\mathcal{W} := V'_{y_1} \cup \cdots \cup V'_{y_n}$ 覆盖. 则 $X \setminus \mathcal{W} \subset W$ 且 $X \setminus \mathcal{W}$ 是包含 $\bigcap_{i=1}^{n} V_{y_i}$ 的闭集合, 因此是 x 的闭邻域.

令 X 是度量空间, V 是 x 的邻域. 则对某一 $\varepsilon > 0$. 开球$B(x,\varepsilon)$ 包含在 V 中. 取闭球$\overline{B}(x,\varepsilon/2)$, 即得到包含在 V 中的闭集.

若 X 是局部凸的拓扑向量空间, 不妨设 $x=0$. 由这样的空间的定义, 0 的邻域 V 包含开的桶集 M. 令 p_M 是与之相联系的 Minkowski 度规, $N := \left\{x \in X : p_M(x) \leqslant \frac{1}{2}\right\}$. 则 N 是 M 中的闭子集, 由引理 2.8, 它是 0 的邻域.

(ii) 由于 S 是闭的, 且 $x \notin S$, 则 $X \setminus S$ 是 x 的开邻域. 由假设, 存在 x 的闭邻域 V 满足 $V \subset X \setminus S$. 则 $W := X \setminus V$ 是包含 S 的开集合, 因此 W 是 S 的邻域. 显然, 集合 V 与 W 的交集是空的. □

多值函数的闭性与上半连续性的某些关系由下述引理表述.

引理 4.3　*设 X 与 U 是 Hausdorff 拓扑空间, 令 $F: U \to 2^X$ 是多值函数. 则下述结论成立:*

(i) 若 X 是紧致的且 F 是闭的, 则 F 是上半连续的.

(ii) 相反地, 若 F 是闭值的且上半连续的, X 的拓扑满足: 一点的邻域包含闭邻域, 则 F 是闭的.

证明　(i) 令 $u_0 \in U, V_X$ 是 $F(u_0)$ 的开邻域. 则 $X \setminus V_X$ 是 X 的闭子集. 因为 X 是紧致的, $X \setminus V_X$ 是紧致的. 考虑点 $x \in X \setminus V_X$. $(u_0, x) \notin \mathrm{gph}(F)$ 是显然的. 由于 $\mathrm{gph}(F)$ 是闭的, 存在 (u_0,x) 的邻域$\mathcal{U}_x \times \mathcal{V}_x$ 满足 $(\mathcal{U}_x \times \mathcal{V}_x) \cap \mathrm{gph}(F) = \varnothing$. 因为 $X \setminus V_X$ 是紧致的, 在 $X \setminus V_X$ 中存在有限多个点 $x_1, \cdots, x_n$ 满足相应的邻域 $\mathcal{V}_{x_i}, i=1,\cdots,n$ 覆盖集合 $X \setminus V_X$, 可取 V_U 是 u_0 的任何包含在相应的邻域 $\mathcal{U}_{x_i}, i=1,\cdots,n$ 的交的邻域.

(ii) 为证明 F 是闭的, 需验证 $\mathrm{gph}(F)$ 在 $U \times X$ 中是闭的, 或等价地, $U \times X \setminus \mathrm{gph}(F)$ 是开的. 考虑点 $(u,x) \notin \mathrm{gph}(F)$, 即 $x \notin F(u)$. 由于 $F(u)$ 是闭的, X 中的任何开邻域均包含闭邻域, 由引理 4.2, 存在 x 的邻域 $\mathcal{V}$ 和开集$\mathcal{W} \subset X$ 满足 $F(u) \subset \mathcal{W}$ 且 $\mathcal{V} \cap \mathcal{W} = \varnothing$. 令 $\mathcal{U}$ 是 u 的邻域, 满足对每一 $u' \in \mathcal{U}$ 有 $F(u') \subset \mathcal{W}$. 则 $\mathcal{U} \times \mathcal{V}$ 是 (u,x) 的邻域, 它与 $\mathrm{gph}(F)$ 的交是空集. 这表明集合 $U \times X \setminus \mathrm{gph}(F)$ 是开的, 因此 F 是闭的. □

命题 4.4　*令 u_0 是参数空间 U 的给定点. 设*

(i) 函数 $f(x,u)$ 在 $X \times U$ 上是连续的.

(ii) 多值函数 $\Phi(\cdot)$ 是闭的.

(iii) 存在 $\alpha \in \mathbb{R}$ 及紧致集合$C \subset X$ 满足在 u_0 的某一邻域中的每一 u, 水平集

$$\mathrm{lev}_\alpha f(\cdot,u) := \{x \in \Phi(u) : f(x,u) \leqslant \alpha\}$$

是非空的且包含在 C 中.

(iv) 对集合 $\mathcal{S}(u_0)$ 的任何邻域 $\mathcal{V}_X$, 存在 u_0 的邻域 $\mathcal{V}_U$ 满足对所有的 $u \in \mathcal{V}_U, \mathcal{V}_X \cap \Phi(u) \neq \varnothing$.

则 (a) 最优值函数 $v(u)$ 在 $u = u_0$ 处是连续的, 且

(b) 多值函数 $u \mapsto \mathcal{S}(u)$ 在 u_0 处是上半连续的.

证明 由假设 (i) 与 (iii), 对 u_0 某一个邻域中的任何 u, 水平集 $\text{lev}_\alpha f(\cdot, u)$ 是非空的闭的, 且包含在紧致集 C 中. 于是有 $\text{lev}_\alpha f(\cdot, u)$ 是紧致的, 因此 $f(\cdot, u)$ 在 $\Phi(u)$ 上取到极小, 即最优解集 $\mathcal{S}(u)$ 是非空的.

令 $\varepsilon > 0$ 是任意正数. 考虑集合

$$S_\varepsilon := \{(u, x) \in \text{gph}(\Phi) : f(x, u) \leqslant v(u_0) - \varepsilon\}.$$

由于集合 $\text{gph}(\Phi)$ 是闭的, $f(x, u)$ 是连续的, 从而得到 S_ε 是闭的. 令 $x \in \Phi(u_0) \cap C$ 满足 $f(x, u_0) > v(u_0) - \varepsilon$. 则 $(u_0, x) \notin S_\varepsilon$, 因此存在 (u_0, x) 的邻域, 它与 S_ε 没有公共点. 因为 $\Phi(u_0) \cap C$ 是紧致的, 则存在 $\Phi(u_0) \cap C$ 与 u_0 的邻域 V_x 与 V_u, 满足对任何 $u \in V_u$ 与 $x \in \Phi(u) \cap V_x$, 有 $f(x, u) > v(u_0) - \varepsilon$. 考虑多值函数 $F(u) := \Phi(u) \cap C$. 由于 Φ 是闭的, C 是紧致的, 引理 4.3 可应用于多值函数 F. 结果, 可以取得邻域 V_u, 满足对所有的 $u \in V_u$, $\Phi(u) \cap C \subset V_x$ 成立. 这就得到对所有的 $u \in V_u$ 与 $x \in \Phi(u) \cap C$, 有 $f(x, u) > v(u_0) - \varepsilon$. 由假设 (iii), 这意味着对所有的 $u \in V_u$, 有 $f(x, u) > v(u_0) - \varepsilon$. 由 ε 的任意性, 可得 $v(\cdot)$ 在 u_0 处是下半连续的.

令 $\varepsilon > 0, \mathcal{V}_X \times \mathcal{V}_U$ 是 $\mathcal{S}(u_0) \times \{u_0\}$ 的邻域满足 $f(x, u) < v(u_0) + \varepsilon$ 对 $(x, u) \in \mathcal{V}_X \times \mathcal{V}_U$ 是成立的. 因为 $\mathcal{S}(u_0)$ 是紧致的, 这样的邻域是存在的, 由假设 (iv), $\mathcal{V}_U$ 可选取满足 $\mathcal{V}_X \cap \Phi(u) \neq \varnothing$, $\forall u \in \mathcal{V}_U$. 这就得到, 对所有的 $u \in \mathcal{V}_U$ 有 $v(u) < v(u_0) + \varepsilon$, 从而 $v(\cdot)$ 在 u_0 处是上半连续的. 得到 $v(\cdot)$ 在 u_0 处是连续的.

由上述讨论, 若 V_X 是 $\mathcal{S}(u_0)$ 的邻域, 则存在某一 $\varepsilon > 0$, 对所有的 u_0 邻域中的 u 与 $x \in \Phi(u) \setminus V_X$ 有 $f(x, u) \geqslant v(u_0) + \varepsilon$. 由 $v(\cdot)$ 的连续性, 证得结论 (b). □

我们讨论当 X 与 U 是 Banach 空间时命题 4.4 的假设. 若 $\Phi(u)$ 由 (4.2) 形式的抽象约束定义, 其中 $G(x, u)$ 是连续的, K 是闭的, 则假设 (ii) 成立. 假设 (iv) 在问题 (P_u) 的框架下与 Robinson 约束规范(2.163) 密切相关. 事实上, 若 (2.163) 在 (x_0, u_0) 处成立, 其中 $x_0 \in \mathcal{S}(u_0)$, 则由定理 2.87 中的稳定性结果 (2.164), 可得

$$\text{dist}(x_0, \Phi(u)) = O(\|G(x_0, u) - G(x_0, u_0)\|), \tag{4.6}$$

因此, 假设 (iv) 由 $G(x, u)$ 的连续性得到. 注意到, 若 $\Phi(u) = \Phi$ 是常值 (与 u 无关) 且集合 Φ 是闭的, 则条件 (ii) 与 (iv) 是自动成立的.

假设 (iii) 称为是下确界紧致条件. 若空间 X 是有限维的向量空间, 则 X 中闭有界子集是紧致的. 此种情况, 不妨假设水平集 $\text{lev}_\alpha f(\cdot, u)$ 是非空的, 且对接近 u_0

的所有的 u 是一致有界的. 对于无穷维空间中的最优化问题, 情况就较难处理. 若 X 是自反的 Banach 空间, 则由 Banach-Alaoglu 定理, X 中有界的弱闭子集是弱紧致的. 进一步, Banach 空间中的凸集合是强闭的当且仅当它是弱闭的, 凸函数是强下半连续当且仅当它是弱下半连续的 (见 2.1.3 节). 所以, 在这样的情况下, 用空间 X 的弱拓扑是合适的.

例 4.5　令 X 是 Hausdorff 拓扑空间, U 是 Banach 空间, $f: X\times U\to\mathbb{R}$. 设 $f(x,u)$ 在 $X\times U$ 上是连续的, 对每一 $x\in X$, $f(x,\cdot)$ 是凸的. 考虑从 $X\times U$ 到 2^{U^*} 的多值函数 $F(x,u):=\partial_u f(x,u)$. 由次微分 $\partial_u f(x,u)$ 的定义及 $f(x,u)$ 的连续性可立即得到多值函数 F 是闭的. 而且, 任何点 $(x_0,u_0)\in X\times U$ 有一邻域使得 $|f(x,u)|$ 在此邻域上是有界的. 结果 (见命题 2.107(v) 证明), $f(x,\cdot)$ 在 u_0 的邻域上的 Lipschitz 常数对 x_0 邻域中所有的 x 是一致有界的. 由命题 2.126(v)(b), $F(x,u)$ 在 (x_0,u_0) 某一邻域中所有的 (x,u) 是一致有界的. 进一步, 设空间 U 是有限维的. 则 U 中的任何有界闭子集均是紧致的, 因此由引理 4.3 可得次微分多值函数 F 是上半连续的.

对于连续性问题, 不同于通常的处理方式是基于增广实值函数上图收敛的概念. 下、上上图极限的定义见 (2.60) 与 (2.61) 及 2.23 节后面的讨论, 增广实值函数序列 $\{f_n\}$ 称为上图收敛到 f, 若序列 $\{f_n\}$ 的下、上上图极限均等于 f. 由 $f_n\xrightarrow{e}f$ 可推出极限函数 f 是下半连续的. 对于我们的分析, 与上图收敛的联系由这样的事实来说明, 即这是具有下述性质的收敛的最弱形式.

命题 4.6　*设 X 是度量空间, $f_n: X\to\overline{\mathbb{R}}$ 是上图收敛到 f 的函数序列. 则*

(i) 下述不等式成立

$$\limsup_{n\to\infty}\left(\inf_{x\in X} f_n(x)\right)\leqslant\inf_{x\in X} f(x). \tag{4.7}$$

(ii) 进一步, 若存在某一序列 $\{n_k\}\subset N, x_k\in\operatorname*{argmin}_{x\in X} f_{n_k}(x)$ 满足 x_k 收敛到点 x^*, 则 $x^*\in\operatorname*{argmin}_{x\in X} f(x)$, 且

$$\lim_{k\to\infty}\left(\inf_{x\in X} f_{n_k}(x)\right)=\inf_{x\in X} f(x). \tag{4.8}$$

证明　令 $x\in X$, $\{x_n\}$ 是与 x 相联系的使 (2.65) 成立的序列. 则

$$\limsup_{n\to\infty}\left(\inf_{x\in X} f_n(x)\right)\leqslant\limsup_{n\to\infty} f_n(x_n)\leqslant f(x).$$

由于上述不等式对所有的 $x\in X$ 均成立, 不等式 (4.7) 成立.

设 $\{n_k\}$ 是上述命题 (ii) 中给出的序列, 令 $\{x_k\}$ 是 $f_{n_k}(\cdot)$ 的极小点序列. 则由 (2.64), $f(x) \leqslant \liminf_{k\to\infty} f_{n_k}(x_{n_k})$①. 结合 (4.7), 即得 (ii) 中的结论. □

把 (P_u) 与下述增广实值函数相联系

$$\bar{f}(x,u) := \begin{cases} f(x,u), & 若x \in \Phi(u), \\ +\infty, & 否则. \end{cases}$$

设 X 与 U 是度量空间, 且

$$\mathrm{e}-\lim_{u\to u_0} \bar{f}(\cdot,u) = \bar{f}(\cdot,u_0). \tag{4.9}$$

由命题 4.6, 若除了 (4.9), 还有最优解集 $\mathcal{S}(u)$ 是非空的且对 u_0 的邻域中的任何 u, u 在 X 中的一紧致子集里, 则命题 4.4 的结论成立.

若函数 $f(x,u)$ 在 $X \times U$ 上是连续的, 则 (4.9) 中的上图极限成立当且仅当 (i) 多值函数 Φ 在 u_0 处是闭的; (ii) 对每一 $\bar{x} \in \Phi(u_0)$ 及任意的 $u_n \to u_0$, 对充分大的 n, 存在 $x_n \in \Phi(u_n)$ 满足 $x_n \to \bar{x}$. 显然, 这样的条件与命题 4.4 中的假设 (ii) 与 (iv) 是密切联系的.

我们简短地讨论上述结果的两个应用. 第一个应用处理类似 3.4.2 节给出的惩罚函数. 考虑在闭集合 $\Phi \subset X$ 上极小化函数 $f: X \to \mathbb{R}$ 的优化问题. 与此问题相联系的是惩罚问题

$$\min_{x\in X} \left\{\phi_r(x) := f(x) + r\psi(x)\right\},$$

其中惩罚函数ψ 是增广的实值函数, 满足对所有的 $x \in X$, $\psi(x) \geqslant 0$, $\psi(x) = 0$ 当且仅当 $x \in \Phi$. 例如, 可像 3.4.2 节那样, 取 $\psi(x) := \operatorname{dist}(x,\Phi)$. 置 $\bar{f}(x) := f(x)$, 若 $x \in \Phi$, 否则, $\bar{f}(x) = +\infty$. 显然, 对所有的 $x \in X$, $\phi_r(x) \leqslant \bar{f}(x)$. 容易验证, 若 f 是下半连续的, 则当 $r \to +\infty$ 时, $\phi_r(x) \xrightarrow{\mathrm{e}} \bar{f}$, 结果, 命题 4.6 的结论是可以应用的.

第二个应用关注 Tichonov 正则化. 即考虑问题

$$\min_{x\in X} \left\{f_r(x) := f(x) + r\|x\|^\alpha\right\},$$

其中 X 是 Banach 空间, f 是 X 上的下半连续的增广实值函数, $\alpha > 0$, $r > 0$. 显然, 当 $r \to 0$ 时, $f_r \xrightarrow{\mathrm{e}} f$.

4.2 方向正则性

这一节讨论 2.3.2 与 2.3.3 节中稳定性分析的方向形式. 对给定的参数空间 U 中的 (方向)d, 考虑具有形式 $u(t) = u_0 + td + o(t)$, $t \in \mathbb{R}_+$ 的路径 $u(t)$. 为推导出

① 应为 $f(x^*) \leqslant \cdots$.

沿路径 $u(t)$ 上最优值函数 $v(u)$ 的上近似, 我们研究构造可行路径 $x(t) \in \Phi(u(t))$ 这一问题. 这一节, 设约束映射 $G(x,u)$ 是 $C^{1,1}$ 的, 即它是可微的且其导数是局部 Lipschitz 连续的. 此种情况下, 由 (2.41) 有 $G(x,u)$ 在点 (x_0,u_0) 处的一阶 Taylor 展式中的余项是 $O(\|x-x_0\|^2+\|u-u_0\|^2)$ 阶的.

定义 4.7 令 $x_0 \in \Phi(u_0)$. 称 $h \in X$ 是在 x_0 处相对于方向 $d \in U$ 的一个一阶可行方向, 若对参数空间 U 中的任何路径$u(t) := u_0+td+o(t)$, 存在 $r(t)=o(t), t \geqslant 0$ 满足 $x_0+th+r(t) \in \Phi(u(t))$.

因为 $G(x_0,u_0) \in K$, 由 Taylor 展式

$$G(x_0+th+o(t), u_0+td+o(t)) = G(x_0,u_0)+tDG(x_0,u_0)(h,d)+o(t),$$

得到 h 是可行方向的必要性条件是

$$\text{dist}\ (G(x_0,u_0)+tDG(x_0,u_0)(h,d),\ K) = o(t),$$

当 K 是凸的时候, 等价于

$$DG(x_0,u_0)(h,d) \in T_K(G(x_0,u_0)). \tag{4.10}$$

若 Robinson 约束规范 (2.163) 成立, 则由稳定性定理 (定理 2.87), 上述条件还是充分的, 此种情况下, 条件 (4.10) 刻画了一阶可行方向.

我们叙述方向正则性条件, 它依赖于扰动方向 d, 在此条件下, (4.10) 刻画了一阶可行方向集合.

定义 4.8 令 $x_0 \in \Phi(u_0)$. 称方向正则性条件在 x_0 沿方向 $d \in U$ 处成立, 若对映射

$$\overline{G}(x,t) := (G(x,u_0+td),t):\ X \times \mathbb{R} \to Y \times \mathbb{R}, \tag{4.11}$$

相对于集合 $K \times \mathbb{R}_+ \subset Y \times \mathbb{R}$, Robinson 约束规范在 $(x_0,0)$ 处成立.

很清楚, 上述方向正则性的概念显式地依赖于所取的方向 d, 而 Robinson 约束规范 (2.163) 并不是这样. 上述定义有些抽象. 下述的刻画在某些特殊情况下理解和验证方向正则性是有用的. $\mathbb{R}_+(d) := \{td : t \geqslant 0\}$.

定理 4.9 下述结论成立:

(i) 方向正则性等价于下述两个条件中的任何一个:

$$0 \in \text{int}\{G(x_0,u_0)+DG(x_0,u_0)(X,\mathbb{R}_+(d))-K\}, \tag{4.12}$$

或存在 $\delta>0$ 与 $\varepsilon>0$ 满足

$$\varepsilon B_Y \subset G(x_0,u_0)+DG(x_0,u_0)(B_X,\delta d)-K. \tag{4.13}$$

(ii) Robinson 约束规范 (2.163) 可推出对任何方向 d 的方向正则性条件.

(iii) 若方向正则性条件成立, 则 $h \in X$ 是可行方向的充分必要条件是它满足 (4.10).

证明 约束映射 $\overline{G}$ 可视为乘积的形式 $\overline{G}(z) = (G_1(z), G_2(z))$, 其中 $z := (x,t) \in X \times \mathbb{R}$, $G_1(z) := G(x, u_0 + td)$, $G_2(z) := t$. 由引理 2.100, 对积映射 $\overline{G}(z)$, 相对于集合 $K \times \mathbb{R}_+$ 的 Robinson 约束规范成立, 当且仅当

$$0 \in \operatorname{int}\{G_1(z_0) + DG_1(z_0)[DG_2(z_0)^{-1}(\mathbb{R}_+ - G_2(z_0))] - K\}, \tag{4.14}$$

其中 $z_0 := (x_0, 0)$①. 由于 $G_2(z_0) = 0$, $DG_2(z_0)(h,t) = t$, 因而有

$$DG_2(z_0)^{-1}(\mathbb{R}_+ - G_2(z_0)) = X \times \mathbb{R}_+,$$

条件 (4.14) 恰好与 (4.12) 相同. 因此, (4.12) 等价于方向正则性条件.

显然, (4.13) 可推出 (4.12). 我们证相反的推出关系. 设 (4.12) 成立. 考虑多值函数

$$\mathcal{M}(h,t) := \begin{cases} G(x_0,u_0) + DG(x_0,u_0)(h,td) - K, & \text{若 } t \geqslant 0, \\ \varnothing, & \text{否则}. \end{cases}$$

这是闭的凸的多值函数, 条件 (4.12) 意味着 $0 \in \operatorname{int}(\operatorname{range}\mathcal{M})$. 由广义的开映射定理 (定理 2.70) 得, 存在 $\varepsilon_1 > 0$,

$$\varepsilon_1 B_Y \subset G(x_0,u_0) + DG(x_0,u_0)(B_X, \Delta(d)) - K, \tag{4.15}$$

其中 $\Delta(d) := \{td : t \in [0,1]\}$.

令 $\varepsilon_2 > 0$ 充分小, 使得 $\varepsilon_2 D_u G(x_0,u_0)d \subset \varepsilon_1 B_Y$. 则由包含关系 (4.15) 可推出

$$-\varepsilon_2 D_u G(x_0,u_0)d \in G(x_0,u_0) + D_x G(x_0,u_0)B_X + \alpha D_u G(x_0,u_0)d - K,$$

对某一 $\alpha \in [0,1]$ 成立. 结果, 对某一 $\varepsilon_3 \geqslant \varepsilon_2 > 0$, 有

$$-\varepsilon_3 D_u G(x_0,u_0)d \in G(x_0,u_0) + D_x G(x_0,u_0)B_X - K,$$

因而有

$$\begin{aligned} -\varepsilon_3 D_u G(x_0,u_0)(\Delta(d)) &\subset \bigcup_{t\in[0,1]} t[G(x_0,u_0) - K] + D_x G(x_0,u_0)B_X \\ &= G(x_0,u_0) - \bigcup_{t\in[0,1]} [(1-t)G(x_0,u_0) + tK] + D_x G(x_0,u_0)B_X \\ &\subset G(x_0,u_0) - K + D_x G(x_0,u_0)B_X. \end{aligned}$$

① 原著中为 $z := (x_0, 0)$.

把此包含关系与 (4.15) 相结合, 可以把它写成下列形式

$$\varepsilon_1 B_Y \subset G(x_0,u_0) - K + DG(x_0,u_0)(B_X,d) - D_uG(x_0,u_0)(\Delta(d)),$$

置 $\tau = 1/\varepsilon_3$, 则可以得到

$$\varepsilon_1 B_Y \subset (1+\tau)(G(x_0,u_0) - K) + DG(x_0,u_0)((1+\tau)B_X,d).$$

将上述包含关系除以 $(1+\tau)$, 令 $\delta := (1+\tau)^{-1}$, $\varepsilon := (1+\tau)^{-1}\varepsilon_1$, 可得 (4.13).

结论 (ii) 可由 (i), 约束正则性的定义及方向正则性的刻画 (4.12) 得到. 现在证 (iii). 我们已经注意到, 任何一阶可行方向满足 (4.10). 相反的结果是下述引理的一个结论. □

引理 4.10　设 $G(x,u)$ 是 $C^{1,1}$ 的, 方向正则性条件在 $x_0 \in \Phi(u_0)$ 处沿方向 d 成立, 令 $u(t) := u_0 + td + o(t)$. 则下述结论成立:

(i) 令 $x(\cdot): \mathbb{R}_+ \to X$ 满足

$$\|x(t) - x_0\| \leqslant \gamma\sqrt{t} \quad 且 \quad \mathrm{dist}\,(G(x(t),u(t)),K) \leqslant \delta t, \tag{4.16}$$

其中 γ,δ 是某一正的常数, $t \geqslant 0$ 充分小. 则对充分小的 γ 与 δ, 存在映射 $\bar{x}(\cdot): \mathbb{R}_+ \to X$ 满足对充分小的 $t \geqslant 0$, 有 $G(\bar{x}(t),u(t)) \in K$, 进一步, 存在某一常数 $c \geqslant 0$,

$$\|\bar{x}(t) - x(t)\| \leqslant c(1 + t^{-1}\|x(t) - x_0\|)\ \mathrm{dist}\,(G(x(t),u(t)),K). \tag{4.17}$$

(ii) 满足 (4.10) 的任何 h 均是一阶可行方向.

证明　结论 (ii) 由 (i) 可得到. 事实上, 若 h 满足 (4.10), 则 $x(t) := x_0 + th$ 满足 (4.16), 其中常数 γ 与 δ 是充分小的. 所以, (4.17) 成立, 存在 $\bar{x}(t) \in \Phi(u(t))$ 满足, 对充分小的 $t \geqslant 0$, 有

$$\|\bar{x}(t) - x(t)\| \leqslant c\,[1 + O(1)]\ \mathrm{dist}\,(G(x(t),u(t)),K) = o(t).$$

剩下要证 (i). 将 (4.13) 与包含关系 $0 \in G(x_0,u_0) - K$ 取权值为 t/δ 及 $(1-t/\delta)$ 的凸组合, 其中 $t \in (0,\delta)$, 得到

$$2t\varepsilon B_Y \subset G(x_0,u_0) + tD_uG(x_0,u_0)d + t\alpha D_xG(x_0,u_0)B_X - K \tag{4.18}$$

对某一 $\varepsilon > 0$ 成立 (此 ε 不同于 (4.13) 中的), 其中 $\alpha := \delta^{-1}$.

我们应用定理 2.84 并用 (4.18). 为了避免混淆, 定理 2.84 中的函数 G 与 H 将用 G' 与 H' 来标记

$$G'(x) := G(x_0,u_0) + DG(x_0,u_0)(x - x_0, td),$$
$$H'(x) := G(x,u(t)).$$

定理 2.84 中的 x_0 取为 $x(t)$, y_0 也记为 y_0, 它满足

$$y_0 \in G'(x(t)) - K, \quad \|y_0\| \leqslant 2 \operatorname{dist}(G'(x(t)), K).$$

将点 (x, y) 取为 $(x(t), 0)$. 由于 $G(x, u)$ 是 $C^{1,1}$ 的, 有

$$\begin{aligned}\|y_0\| &\leqslant 2\|G'(x(t)) - G(x(t), u(t))\| + 2\operatorname{dist}(G(x(t), u(t)), K)\\ &= O(\|x(t) - x_0\|^2 + t^2) + 2\delta t\\ &= O(\gamma^2 + \delta)t \leqslant \varepsilon t/10,\end{aligned}$$

其中最后的不等式对不依赖于 t 的充分小的 γ 与 δ 是成立的. 由 (4.18), 对充分小的 $t > 0$ 有

$$2t\varepsilon B_Y \subset G'(x(t)) + (\alpha t + \|(x(t) - x_0)\|)D_xG(x_0, u_0)B_X - K. \tag{4.19}$$

由于 $\|y_0\| \leqslant \varepsilon t/10$, 可推出

$$y_0 + t\varepsilon B_Y \subset \mathcal{F}_{G'}(x(t) + (\alpha t + \|x(t) - x_0\|)B_X)^{①}$$

令

$$\eta := t\varepsilon, \quad \upsilon := \alpha t + \|x(t) - x_0\| \leqslant \alpha t + \gamma\sqrt{t} < 2\gamma\sqrt{t},$$

对充分小的 $t > 0$, 条件 (2.130) 是成立的. 由定理 2.83 得, 若

$$\|x - x(t)\| < \frac{1}{2}\upsilon \leqslant \gamma\sqrt{t}, \quad \|y - y_0\| < \frac{1}{8}\eta = \frac{1}{8}t\varepsilon, \tag{4.20}$$

则度量正则性条件 (2.133) 成立, 其中

$$c = \frac{4\upsilon}{\eta} = \frac{4(\alpha t + \|x(t) - x_0\|)}{t\varepsilon} < 8\gamma\varepsilon^{-1}t^{-1/2}.$$

注意到, 对充分小的 γ 与 δ, 因为 $x = x(t)$, $y = 0$, $\|y_0\| \leqslant t\varepsilon/10$, 所以条件 (4.20) 成立.

现在应用定理 2.84, 有

$$\eta_x = \frac{1}{2}(\alpha t + \|x(t) - x_0\|) < \gamma\sqrt{t}, \quad \eta_y = \frac{1}{8}t\varepsilon.$$

需要对满足 (2.160) 和 (2.161) 的常数 η_x^1 与 η_y^1 检验 (2.147) 是成立的. 因为 $DG'(\cdot) = DG(x_0, u_0)$, $DH'(\cdot) = DG(\cdot, u(t))$, 在 $B_X(x(t), \eta_x)$ 中, $D(x) = G(x) - H(x)$ 的 Lipschitz 常数 κ 满足 $\kappa = O(\gamma)\sqrt{t}$. 所以 ($\varepsilon$ 是固定的)$c\kappa \leqslant O(\gamma^2)$ 对充分小的 γ 是小于 1/2 的, 则

$$c(\kappa) = c(1 - c\kappa)^{-1} \leqslant \frac{8(\alpha t + \|x(t) - x_0\|)}{t\varepsilon} = O(1 + t^{-1}\|x(t) - x_0\|).$$

① 应为 $y_0 + t\varepsilon B_Y \subset \mathcal{F}_{G'}(x(t)) + (\alpha t + \|x(t) - x_0\|)D_xG(x_0, u_0)B_X$.

因为 $x - x_0 = 0$ 且

$$\|y - (y_0 - D(x_0))\| = \|y_0 - D(x_0)\| \leqslant \|y_0\| + \|D(x_0)\| = O(\gamma^2 + \delta)t,$$

可以取 η_x^1 任意的小, 通过取 γ 与 δ 充分接近于 0, 可设对任意小 $\theta > 0$, 对充分小的 $t > 0$, 有 $\eta_y^1 \leqslant \theta t$. 则 (2.160) 和 (2.161) 成立. 由定理 2.84 得, 存在 $\bar{x}(t)$ 满足引理的结论. □

我们给出一些注记来结束本节的讨论. 方向正则性条件 (4.12) 不能推出 Lagrange 乘子的存在性. 类似于 2.3.4 节中的讨论, 当约束具有特殊形式时, 方向正则性条件具有有用的刻画. 尤其, 当 X 是有限维时, 对有限约束问题, 方向正则性等价于 Gollan 条件 (与 (2.191) 比较)

$$\begin{aligned}
&Dg_i(x_0, u_0), \quad i = 1, \cdots, q \text{ 是线性无关的,} \\
&\exists h \in X : Dg_i(x_0, u_0)(h, d) = 0, \quad i = 1, \cdots, q, \\
&\qquad Dg_i(x_0, u_0)(h, d) < 0, \quad \forall i \in I(x_0, u_0),
\end{aligned} \tag{4.21}$$

其中

$$I(x_0, u_0) := \{i : g_i(x_0, u_0) = 0,\ i = q + 1, \cdots, p\}.$$

约束在乘积空间 $Y = Y_1 \times Y_2$ 的情况, 其中 $K = \{0\} \times K_2$, K_2 具有非空内部, 这一刻画是 (与推论 2.101 相比较)

$$\begin{aligned}
&D_x G_1(x_0, u_0) \text{ 是映上的,} \\
&\exists h \in X,\ \exists \varepsilon > 0 : DG_1(x_0, u_0)(h, d) = 0, \\
&\qquad G_2(x_0, u_0) + DG_2(x_0, u_0)(h, \varepsilon d) \in \operatorname{int}(K_2).
\end{aligned} \tag{4.22}$$

当问题具有下述形式

$$\min_x f(x, u) \quad \text{s. t.} \quad x \in \mathcal{C}, \quad \mathcal{G}(x, u) \in \mathcal{K},$$

其中, $\mathcal{C}$ 与 $\mathcal{K}$ 是闭凸集, f 与 $\mathcal{G}$ 是光滑映射. $\mathcal{K}$ 具有非空的内部, 有如下刻画

$$\exists x' \in \mathcal{C},\ \exists \varepsilon > 0 : \mathcal{G}(x_0, u_0) + D\mathcal{G}(x_0, u_0)(x' - x_0, \varepsilon d) \in \operatorname{int}\ (\mathcal{K}). \tag{4.23}$$

设 $\mathcal{D}$ 是方向正则性成立的方向 $d \in U$ 的集合. 在引理 4.10 的证明中已证得方向正则性可推出 (4.18). 由于相反的结论是显然的, (4.18) 刻画了方向正则性成立的方向 d. 由 (4.18) 得 $\mathcal{D}$ 是凸锥. 由于方向正则性是辅助问题的正则性条件, (由定理 2.84) 则它在小的扰动下是稳定的, 得到 $\mathcal{D}$ 是开凸锥. 称 $\mathcal{D}$ 为正则方向锥. 注意到锥 $\mathcal{D}$ 可能是空集. 还注意到, Robinson 约束规范等价于 $d = 0$ 时的方向正则性. 所

以, 若 $0 \in \mathcal{D}$, 则 $\mathcal{D} = U$. 结果, 若对某一 $d \neq 0$ 方向正则性条件成立, $-d$ 的方向正则性也成立 (即若正则方向锥不是点的), 则 Robinson 约束规范成立, 因此 $\mathcal{D} = U$.

由 (4.12) 得, 若正则方向锥 $\mathcal{D}$ 非空, 则下述条件成立

$$0 \in \operatorname{int}\{G(x_0, u_0) + DG(x_0, u_0)(X \times U) - K\}. \tag{4.24}$$

此条件可被认为是映射 $G(x, u)$ 以 x 与 u 为联合变量的 Robinson 约束规范. 相反地, 若集合 K 具有非空的内部, 则由 (2.186), 有 d 是正则性方向当且仅当存在 $h \in X$ 满足

$$DG(x_0, u_0)(h, d) \in \operatorname{int}(T_K(G(x_0, u_0))). \tag{4.25}$$

因此, 此种情形下, (4.24) 可推出正则方向的存在性.

4.3 最优值函数的一阶可微性分析

这一节讨论最优值函数 $v(u)$ 在给定点 $u_0 \in U$ 处的一阶可微性性质. 我们通过考虑可行集是非扰动 (即不依赖于 u 的情况) 来开始分析.

4.3.1 固定的可行集的情况

这一节讨论可行集 $\Phi(u)$ 不依赖于 u 的情况, 即对所有的 $u \in U$, $\Phi(u) = \Phi$, 从而 $v(u)$ 由下述参数化问题的最优值函数给出

$$(\mathrm{P}_u) \qquad \min_{x \in X} f(x, u) \qquad \text{s. t.} \qquad x \in \Phi. \tag{4.26}$$

除非特别说明, 这一节假设 U 是 Banach 空间, X 是 Hausdorff 拓扑空间, $\Phi \subset X$ 是非空的闭的, $f : X \times U \to \mathbb{R}$ 是连续的.

例 4.11 令 $f(x, u) := xu$, x, $u \in \mathbb{R}$, $\Phi := [-1, 1]$. 容易验证 $v(u) = -|u|$. 尽管这里的可行集是非扰动的且是紧致的, 最优值函数在 $u = 0$ 处不是可微的.

这个例子说明, 始于光滑的数据会终止于不可微的最优值函数. 尽管 $v(u)$ 在 0 处是不可微的, 但沿正的与负的方向, 它均具有方向导数. 这样的方向可微性质对于最优值函数而言是典型的. 在固定的可行集的情况, 对最优值函数的一阶可微性质可给出相当完整的刻画. 下确界紧致性条件成立, 若存在 $\alpha \in \mathbb{R}$, 紧致集合 $C \subset X$ 满足: 对接近于 u_0 的每一 u, 水平集

$$\operatorname{lev}_\alpha f(\cdot, u) := \{x \in \Phi : f(x, u) \leqslant \alpha\}$$

是非空的且包含在集合 C 中.

命题 4.12 设 (i) 函数 $f(x, u)$ 在 $X \times U$ 上是连续的.

(ii) 下确界紧致性条件成立.

(iii) 对每一 $x \in \Phi$, 函数 $f_x(\cdot) := f(x, \cdot)$ 在 u_0 处是方向可微的.

(iv) 若 $d \in U$, $t_n \downarrow 0$, $\{x_n\}$ 是 C 中的序列, 则 $\{x_n\}$ 有极限点 $\bar{x}$ 满足

$$\limsup_{n\to\infty} \frac{f(x_n, u_0 + t_n d) - f(x_n, u_0)}{t_n} \geqslant f'_{\bar{x}}(u_0, d). \tag{4.27}$$

则最优值函数 $v(u)$ 在 u_0 处是方向可微的, 且

$$v'(u_0, d) = \inf_{x\in\mathcal{S}(u_0)} f'_x(u_0, d). \tag{4.28}$$

进一步, 对某一序列 $t_n \downarrow 0$, $x_n \in \mathcal{S}(x_0 + t_n d)$, 则 $\{x_n\}$ 中的任何极限点 $\bar{x}$ 属于 $\mathcal{S}_1(u_0, d)$①, 其中

$$\mathcal{S}_1(u_0, d) := \operatorname*{argmin}_{x\in\mathcal{S}(u_0)} f'_x(u_0, d). \tag{4.29}$$

证明 由上述的假设 (i) 与 (ii), 集合 $\mathcal{S}(u_0)$ 是非空的. 固定 $\bar{x} \in \mathcal{S}(u_0)$, $d \in U$. 由 $f(\bar{x}, \cdot)$ 在 u_0 处的方向可微性, 对于 $t \geqslant 0$ 有

$$f(\bar{x}, u_0 + td) - f(\bar{x}, u_0) = t f'_{\bar{x}}(u_0, d) + o(t). \tag{4.30}$$

由于 $v(u_0 + td) \leqslant f(\bar{x},\ u_0 + td)$, $v(u_0) = f(\bar{x}, u_0)$, 由 (4.30) 可得

$$\limsup_{t\downarrow 0} \frac{v(u_0 + td) - v(u_0)}{t} \leqslant f'_{\bar{x}}(u_0, d).$$

对 $\bar{x} \in \mathcal{S}(u_0)$ 取极小, 得

$$v'_+(u_0, d) \leqslant \inf_{x\in\mathcal{S}(u_0)} f'_x(u_0, d). \tag{4.31}$$

为了证明另一方向的不等式, 令 $t_n \downarrow 0$ 满足

$$v'_-(u_0, d) = \lim_{n\to\infty} \frac{v(u_0 + t_n d) - v(u_0)}{t_n}.$$

由上述假设 (i) 与 (ii), 对充分大的 n, 最优解集 $\mathcal{S}(u_0 + t_n d)$ 是非空的, 因此存在 $x_n \in \mathcal{S}(u_0 + t_n d)$. 由紧致性假设, 它有极限点 $\bar{x}$ 满足 (4.27). 由命题 4.4 有 $\bar{x} \in \mathcal{S}(u_0)$. 进一步, 由假设 (iv), 即 (4.27), 因为 $v(u_0 + t_n d) = f(x_n, u_0 + t_n d)$ 与 $v(u_0) \leqslant f(x_n, u_0)$, 可得

$$v'_-(u_0, d) = \lim_{n\to\infty} \frac{v(u_0 + t_n d) - v(u_0)}{t_n} \geqslant f'_{\bar{x}}(u_0, d). \tag{4.32}$$

① 此 x_n 即 (iv) 中的 x_n.

结合 (4.31), 可推出 $v(\cdot)$ 在 u_0 处是方向可微的且 (4.28) 成立.

由 (4.31) 和 (4.32) 可得 $\bar{x} \in \mathcal{S}_1(u_0, d)$, 证毕. □

我们作出下述的观察. 上界 (4.31) 不涉及上述命题中的假设 (iv). 然而, 为导出下界 (4.32), 假设 (iv) 是本质的. 若除了上述命题的假设, 还假设 $f(x,\cdot)$ 是 Hadamard 方向可微的, 且将 (4.27) 中的 d 用 $d_n \to d$ 来代替时亦是成立的, 则 $v(\cdot)$ 亦是 Hadamard 方向可微的.

有两种重要的情形, 命题 4.12 中的假设 (iv) 是自然成立的. 下述结果本质上归功于 Danskin, 可视为有限多个函数极大值的方向导数公式 (2.124) 的一个推广.

定理 4.13 设对所有的 $x \in X$, 函数 $f(x,\cdot)$ 是 (Gâteaux) 可微的, $f(x,u)$ 与 $D_u f(x,u)$ 在 $X \times U$ 上是连续的, 下确界紧致性条件成立. 则最优值函数 $v(\cdot)$ 在 u_0 处是 Fréchet 方向可微的, 且

$$v'(u_0, d) = \inf_{x \in \mathcal{S}(u_0)} D_u f(x, u_0) d. \tag{4.33}$$

进一步, 若对 $t_n \downarrow 0$, $d_n \to d$, $x_n \in \mathcal{S}(u_0 + t_n d_n)$, 则由 $x_n \to \bar{x}$ 可得 $\bar{x} \in \mathcal{S}_1(u_0, d)$, 其中

$$\mathcal{S}_1(u_0, d) := \operatorname*{argmin}_{x \in \mathcal{S}(u_0)} D_u f(x, u_0) d. \tag{4.34}$$

证明 为证明 $v(u)$ 的方向可微性, 只需要证明命题 4.12 中的假设 (iv). 然而, 因为我们要证明 $v(u)$ 是 Fréchet 方向可微的, 直接导出 (4.33) 将更加容易. 注意到, 因为假设 $D_u f(x,u)$ 是连续的, 因此 $f(x,\cdot)$ 是连续可微的, 于是 $f(x,\cdot)$ 是 Fréchet 可微的.

令 $\bar{x} \in \mathcal{S}(u_0)$. 因为 $v(u_0) = f(\bar{x}, u_0)$, $v(u_0 + d) \leqslant f(\bar{x}, u_0 + d)$, 有

$$v(u_0 + d) - v(u_0) \leqslant f(\bar{x}, u_0 + d) - f(\bar{x}, u_0).$$

进一步, 由中值定理

$$\|f(\bar{x}, u_0 + d) - f(\bar{x}, u_0) - D_u f(\bar{x}, u_0) d\| \leqslant \|d\| \int_0^1 \|D_u f(\bar{x}, u_0 + td) - D_u f(\bar{x}, u_0)\| \mathrm{d}t.$$

置

$$\phi(x, u) := \|D_u f(x, u) - D_u f(x, u_0)\|.$$

由于 $\phi(x,u)$ 是连续函数, $\phi(x, u_0) = 0$, $\forall x$, $\mathcal{S}(u_0)$ 是紧致的, 有 $\phi(x,u)$ 关于 $x \in \mathcal{S}(u_0)$ 为一致的收敛到 0, 当 $u \to u_0$ 时. 从而有

$$\|f(\bar{x}, u_0 + d) - f(\bar{x}, u_0) - D_u f(\bar{x}, u_0) d\| = o(\|d\|), \tag{4.35}$$

且 (4.35) 的右端是对 $\bar{x} \in \mathcal{S}(u_0)$ 为一致的 $o(\|d\|)$ 阶的. 结果有

$$v(u_0 + d) - v(u_0) \leqslant \inf_{x \in \mathcal{S}(u_0)} D_u f(x, u_0)d + o(\|d\|). \tag{4.36}$$

相反地, 由 (4.36), 只需证明, 对每一 $\varepsilon > 0$, 对接近于 0 的 d, 有

$$v(u_0 + d) - v(u_0) \geqslant \inf_{x \in \mathcal{S}(u_0)} D_u f(x, u_0)d - \varepsilon(\|d\|).$$

假设存在 $\varepsilon > 0$, 这对收敛于 $0 \in U$ 的序列 $\{d_n\}$ 是不成立的. 由下确界紧致性假设, 最优解集 $\mathcal{S}(u_0 + d_n)$ 是非空的, 因而存在点 $x_n \in \mathcal{S}(u_0 + d_n)$. 有

$$v(u_0 + d) - v(u_0) \geqslant f(x_n, u_0 + d_n) - f(x_n, u_0).$$

由下确界紧致性假设, 序列 $\{x_n\}$ 有极限点 $\bar{x}$, 且 $\bar{x} \in \mathcal{S}(u_0)$. 令 $\mathcal{O}$ 是 $(\bar{x}, u_0)$ 的邻域满足 $\phi(x, u) \leqslant \dfrac{1}{2}\varepsilon$, $\forall (x, u) \in \mathcal{O}$. 则对某一子序列 $n(k)$, 满足 $(x_{n(k)}, u_{n(k)}) \in \mathcal{O}$, 因此由中值定理有

$$\begin{aligned} &\|f(x_n, u_0 + d_n) - f(x_n, u_0) - D_u f(x_n, u_0)d_n\| \\ \leqslant\ &\|d_n\| \, \| \textstyle\int_0^1 D_u f(x_n, u_0 + td_n) - D_u f(x_n, u_0)\| dt \leqslant \dfrac{1}{2}\varepsilon\|d_n\|. \end{aligned}$$

结果

$$\begin{aligned} v(u_0 + d_n) - v(u_0) &\geqslant D_u f(\bar{x}, u_0)d_n - \frac{1}{2}\varepsilon\|d_n\| \\ &\geqslant \inf_{x \in \mathcal{S}(u_0)} D_u f(x, u_0)d_n - \frac{1}{2}\varepsilon\|d_n\|, \end{aligned}$$

这与假设矛盾, 所以, (4.33) 与 Fréchet 意义下的 $v(u)$ 的方向可微性是成立的. 进一步, 若 $d_n/\|d_n\| \to d$, 则 $\bar{x} \in \mathcal{S}_1(u_0, d)$, 这完成了证明. □

注 4.14 注意, 因为在下确界紧致性假设之下, 集合 $\mathcal{S}(u_0)$ 是紧致的, 由 $D_u f(\cdot, u_0)d$ 的连续性得, (4.33) 中的下确界是可以取到的. 在上述定理的假设下, 最优值函数在 u_0 处是 Fréchet 可微的充分必要条件是 $D_u f(x, u_0)$ 对所有的 $x \in \mathcal{S}(u_0)$ 是相同的. 尤其, 若最优解集 $\mathcal{S}(u_0) = \{\bar{x}\}$ 是单点集, $v(\cdot)$ 在 u_0 处是 Fréchet 可微的. 此种情形, $Dv(u_0) = D_u f(\bar{x}, u_0)$.

注 4.15 考虑最大值函数

$$\bar{v}(u) := \sup_{x \in \Phi} f(x, u). \tag{4.37}$$

设上述定理 4.13 中的假设成立, 其中下确界紧致性条件被替换为相应的上确界紧致性条件. 则可得到

$$\bar{v}'(u_0, d) = \sup_{x \in \bar{\mathcal{S}}(u_0)} D_u f(x, u_0)d, \tag{4.38}$$

其中 $\bar{S}(u) := \underset{x\in\Phi}{\operatorname{argmax}} f(x,u)$. 即 $\bar{v}'(u_0,\cdot)$ 是凸的正齐次函数, 它在 $d=0$ 的次微分由集合

$$\operatorname{clconv}\Big(\bigcup_{x\in\bar{S}(u_0)} D_u f(x,u_0)\Big) \tag{4.39}$$

给出. 注意到上述集合是闭的. 事实上, 由上确界紧致性条件可得 $\bar{S}(u_0)$ 是紧致的, 结果, 由于 $D_u f(\cdot,u_0)$ 是连续的, (4.39) 中的括号内的集合是紧致的, 因此它的凸包是紧致的, 因而是闭的.

命题 4.12 中的假设 (iv) 可以被验证的另一种情形是下述定理中的凸的 (或凹的) 情况.

定理 4.16 *设对所有的 $x\in X$, 函数 $f(x,\cdot)$ 是凹的, $f(x,u)$ 在 $X\times U$ 上是连续的, 下确界紧致性条件成立. 则最优值函数 $v(\cdot)$ 在 u_0 处是方向可微的, 且公式 (4.28) 成立.*

证明 由于 $f(x,\cdot)$ 是凹的 (即 $-f(x,\cdot)$ 是凸的) 且是连续的, 则 $f(x,\cdot)$ 是方向可微的 (见 2.4.3 节). 所以我们只需要验证命题 4.12 中的假设 (iv) 成立.

用反证法. 假设 (4.27) 对某一序列 $t_n\downarrow 0$ 与 $x_n\to\bar{x}$ 不真. 则存在 $\varepsilon>0$, 满足

$$\frac{f(x_n,u_0+t_nd)-f(x_n,u_0)}{t_n} \leqslant f'_{\bar{x}}(u_0,d)-\varepsilon.$$

因为 $f(\bar{x},\cdot)$ 是方向可微的, 对某一 $\bar{t}>0$ 有

$$\frac{f(\bar{x},u_0+\bar{t}d)-f(\bar{x},u_0)}{\bar{t}} \geqslant f'_{\bar{x}}(u_0,d)-\frac{1}{2}\varepsilon.$$

由上述两个不等式可得

$$\frac{f(x_n,u_0+t_nd)-f(x_n,u_0)}{t_n}+\frac{1}{2}\varepsilon \leqslant \frac{f(\bar{x},u_0+\bar{t}d)-f(\bar{x},u_0)}{\bar{t}}. \tag{4.40}$$

由 $f(x_n,\cdot)$ 是凹的, 有 (4.40) 的左端的比例率是随 t_n 递减到 0 而单调递增的. 所以, 对 $t_n\leqslant\bar{t}$, 由 (4.40) 得

$$\frac{f(x_n,u_0+\bar{t}d)-f(x_n,u_0)}{\bar{t}}+\frac{1}{2}\varepsilon \leqslant \frac{f(\bar{x},u_0+\bar{t}d)-f(\bar{x},u_0)}{\bar{t}}. \tag{4.41}$$

当 $n\to\infty$ 时, 在 (4.41) 中取极限, 由 $f(x,u)$ 的连续性, 可得到矛盾. □

当然, 上述定理也可对最大值函数 (4.37) 进行描述. 由对偶性, 方向导数的相应的公式可翻译为次微分的公式.

定理 4.17　设对所有的 $x \in X$, 函数 $f(x,\cdot)$ 是凸的, $f(x,u)$ 在 $X \times U$ 上是连续的, 且上确界紧致性条件成立. 则最大值函数 $\bar{v}(u)$ 是凸的连续的, 且

$$\partial\bar{v}(u_0) = \operatorname{clconv}\Big(\bigcup_{x\in\bar{S}(u_0)} \partial_u f(x,u_0)\Big). \tag{4.42}$$

证明　因为对所有的 x, $f(x,\cdot)$ 是凸的, 则最大值函数 $\bar{v}(\cdot)$ 也是凸的. 由命题 4.4, $\bar{v}(\cdot)$ 是连续的. 由定理 4.16 得

$$\bar{v}'(u_0,d) = \sup_{x\in\bar{S}(u_0)} f'_x(u_0,d), \tag{4.43}$$

其中 $f_x(\cdot) := f(x,\cdot)$. 由于 $\bar{v}(\cdot)$ 与 $f_x(\cdot)$ 是连续的, 还有 $\bar{v}'(u_0,\cdot)$ 与 $f'_x(u_0,\cdot)$ 分别是集合 $\partial\bar{v}(u_0)$ 与 $\partial_u f(x,u_0)$ 的支撑函数. 由命题 2.116 关于支撑函数的公式 (2.223), 由 (4.43) 得 (4.42) 中左端集合与右端集合的支撑函数是重合的. 由于这两个集合是闭凸的, 它们是相等的. 这完成了证明. □

注 4.18　若空间 U 是有限维的, 上述定理的假设成立, 则集合 $\bigcup_{x\in\bar{S}(u_0)} \partial_u f(x,u_0)$ 是紧致的. 事实上, 因为对每一 $x \in \bar{S}(u_0)$, 次微分 $\partial_u f(x,u_0)$ 是有界的, $\bar{S}(u_0)$ 是紧致的, 由多值函数 $x \mapsto \partial f_x(u_0)$ 的上半连续性 (见例 4.5), 这一集合是有界的且是闭的. 此集合的凸包也是紧致的, 从而没有必要在公式 (4.42) 的右端取拓扑包.

例 4.19　令 $C(\Omega)$ 是紧致度量空间 Ω 上的连续函数空间. 考虑最大值函数 $\psi : C(\Omega) \to \mathbb{R}$, 定义为 $\psi(x) := \sup_{\omega\in\Omega} x(\omega)$. 这是凸函数, 可视它为最优值函数, 其中 $X := \Omega$, $U := C(\Omega)$. 由定理 4.13 可得, $\psi(\cdot)$ 是 Fréchet 方向可微的, 且

$$\psi'(x,h) = \sup_{\omega\in\Theta(x)} h(\omega), \tag{4.44}$$

其中 $\Theta(x) := \operatorname{argmax}_{\omega\in\Omega} x(\omega)$. 公式 (4.44) 也可以通过关于 $\partial\psi(x)$ 的公式 (2.236) 推导出来.

例 4.20　考虑 $p\times p$ 矩阵空间 $\mathcal{S}^p$, $\lambda_{\max}(A)$ 记矩阵 $A \in \mathcal{S}^p$ 的最大特征值 (见例 2.65). 有

$$\lambda_{\max}(A) = \max_{\|x\|=1} x^{\mathrm{T}}Ax,$$

上面的最大值在由具有单位长度的 A 的特征向量形成的集合 $\mathcal{E}$ 上达到, 特征向量对应着最大特征值 $\lambda_{\max}(A)$. 因为 $x^{\mathrm{T}}Ax = \operatorname{trace} Axx^{\mathrm{T}}$, 由 Danskin 定理4.13 得

$$\lambda'_{\max}(A,H) = \max_{x\in\mathcal{E}}\{\operatorname{trace} Hxx^{\mathrm{T}}\}.$$

令 s 是最大特征值 $\lambda_{\max}(A)$ 的重数, $E := [e_1, \cdots, e_s]$ 是 $p\times s$ 矩阵, 它的列 $e_1, \cdots, e_s$ 构成了对应于最大特征值的 A 的特征向量空间的一组正交基. 有

$$\mathcal{E} = \{x = E\alpha : \ \alpha \in \mathbb{R}^s, \ \|\alpha\| = 1\},$$

从而

$$\lambda'_{\max}(A, H) = \max_{\|\alpha\|=1} \alpha^{\mathrm{T}} E^{\mathrm{T}} H E \alpha = \lambda_{\max}(E^{\mathrm{T}} H E).$$

4.3.2 在抽象约束下的最优值函数的方向可微性

研究可行集也依赖于扰动变量最优值函数的微分性质是相当困难的. 我们回到由 (4.1) 给出的一般问题 (P_u), 其中可行集由 (4.2) 形式的抽象约束定义. 这一节, 设 X, Y 与 U 是 Banach 空间, $f(x,u)$ 与 $G(x,u)$ 是连续可微的. 令 $x_0 \in \mathcal{S}(u_0)$. 对给定的方向 $d \in U$, 考虑 (P_u) 的线性化:

$$(\mathrm{PL}_d) \qquad \min_{h\in X} Df(x_0,u_0)(h,d) \quad \text{s. t.} \quad DG(x_0,u_0)(h,d) \in T_K(y_0), \tag{4.45}$$

其中 $y_0 := G(x_0, u_0)$.

问题 (PL_d) 是锥线性问题. 在 2.5.6 节已经讨论过这样的锥线性问题. 不难计算这一问题的 (Lagrange) 对偶. 对偶问题的有效域由满足下述方程的 $\lambda \in [T_K(y_0)]^-$ 组成

$$D_x f(x_0,u_0) + [D_x G(x_0,u_0)]^* \lambda = 0.$$

由于 $[T_K(y_0)]^- = N_K(y_0)$, 对偶问题的有效域恰好是 $\Lambda(x_0,u_0)$, 它是问题 (P_{u_0}) 在 x_0 处的 Lagrange 乘子集合. 因此 (PL_d) 的对偶可以表示为

$$(\mathrm{DL}_d) \qquad \min_{\lambda\in\Lambda(x_0,u_0)} D_u L(x_0,\lambda,u_0)d. \tag{4.46}$$

由命题 2.191, 若 $\bar{h}$ 与 $\bar{\lambda}$ 分别是 (PL_d) 与 (DL_d) 的可行点, 则 $\bar{h}$ 与 $\bar{\lambda}$ 分别是 (PL_d) 与 (DL_d) 的最优解. (PL_d) 与 (DL_d) 间没有对偶间隙的充分必要条件是

$$\langle \bar{\lambda}, DG(x_0,u_0)(\bar{h},d)\rangle = 0. \tag{4.47}$$

问题 (PL_d) 的 Robinson 约束规范可以写为

$$0 \in \operatorname{int}\{D_x G(x_0,u_0)X + D_u G(x_0,u_0)d - T_K(y_0)\}. \tag{4.48}$$

另一方面, 由 (4.13), 方向正则性条件可推出

$$0 \in \operatorname{int}\{D_x G(x_0,u_0)X + D_u G(x_0,u_0)d - \mathcal{R}_K(y_0)\}. \tag{4.49}$$

显然, (4.48) 可由 (4.49) 得到, 因此方向正则性条件可推出问题 (PL_d) 的 Robinson 约束规范. 下述结果是定理 2.165 的一个结论.

命题 4.21 设方向正则性条件 (4.12) 成立. 则 $\text{val}(PL_d) = \text{val}(DL_d) < +\infty$, 其公共值 $\text{val}(PL_d) = \text{val}(DL_d)$ 是有限的当且仅当 Lagrange 乘子集合 $\Lambda(x_0, u_0)$ 是非空的. 此种情形, 对偶问题(DL_d) 的最优解集合是 $\Lambda(x_0, u_0)$ 的非空的有界弱 * 紧致子集.

注意到, 在命题 4.21 的前提下, 可能发生即使 $\text{val}(PL_d)$ 是有限的, 但问题 (PL_d) 没有最优解的情况. 命题 4.50 将给出 (PL_d) 的解存在的某些充分条件. 若可行对 $(\bar{h}, \bar{\lambda})$ 满足 (4.47), 则解集 $\mathcal{S}(PL_d)$ 的回收锥与临界锥$C(x_0)$ 重合. 事实上, 此种情况有 $\bar{\lambda} \in \mathcal{S}(DL_d)$, $\bar{h} \in \mathcal{S}(PL_d)$, 在 ($PL_d$) 与 ($DL_d$) 间不存在对偶间隙, h 是 (PL_d) 的最优解的充分必要条件是它是可行的且互补条件成立, 即

$$\mathcal{S}(PL_d) = \{h : DG(x_0, u_0)(h, d) \in T_K(y_0),\ \langle \bar{\lambda}, DG(x_0, u_0)(h, d)\rangle = 0\}. \tag{4.50}$$

点 h 属于 $\mathcal{S}(PL_d)$ 的回收锥的充分必要条件是对任何 $t \geqslant 0$, 点 $\bar{h} + th$ 对于 (PL_d) 是可行的且满足互补条件 (4.47). 由 (3.20), 这一条件成立的充要条件是 $h \in C(x_0)$. 还注意到, 若约束映射 G 不依赖于 u 且 $\Lambda(x_0, u_0) \neq \varnothing$, 则 $\mathcal{S}(PL_d)$ 等于 $C(x_0)$, 因而是非空的.

命题 4.22 令 $x_0 \in \mathcal{S}(u_0)$ 满足, 方向正则性条件 (4.12) 在方向 $d \in U$ 处成立. 则对任何 $u(t) := u_0 + td + o(t)$,

$$\limsup_{t \downarrow 0} \frac{v(u(t)) - v(u_0)}{t} \leqslant \text{val}(PL_d). \tag{4.51}$$

证明 令 h 是问题 (PL_d) 的可行点. 则有 $DG(x_0, u_0)(h, d) \in T_K(y_0)$. 由引理 4.10, 在方向正则性条件下, h 是一可行方向①. 即存在 $r(t) = o(t)$, $t \geqslant 0$ 满足 $x_0 + th + r(t) \in \Phi(u(t))$. 则

$$\begin{aligned} v(u(t)) &\leqslant f(x_0 + th + r(t), u(t)) \\ &= f(x_0, u_0) + tDf(x_0, u_0)(h, d) + o(t), \end{aligned}$$

因此, 由 $v(u_0) = f(x_0, u_0)$ 可得到

$$\limsup_{t \downarrow 0} \frac{v(u(t)) - v(u_0)}{t} \leqslant Df(x_0, u_0)(h, d).$$

因为 h 是 (PL_d) 的任意可行点, 得到 (4.51), 这就完成了证明. □

上述结果表明, 在方向正则性条件下, $v'_+(u_0, d) \leqslant \text{val}(PL_d)$, 其中上方向导数 $v'_+(u_0, d)$ 可以按 Hadamard 意义来考虑.

① 相对于 d.

当然可以把命题 4.21 与 4.22 的结果结合起来. 尤其, 若方向正则性条件 (4.12) 成立且 Lagrange 乘子集合 $\Lambda(x_0, u_0)$ 是空集, 则 $v'(u_0, d) = -\infty$, 其中方向导数 $v'(u_0, d)$ 可以按 Hadamard 定义选取. 4.8.3 节将讨论这种 Lagrange 乘子不存在的情况.

一个自然的问题是, (4.51) 给出, (当 $\mathcal{S}(u_0) = \{x_0\}$ 时) 的上界是否在下述定义下是紧的, 即 $v(u)$ 在 u_0 处沿方向 d 是 (Hadamard) 方向可微的且方向导数等于 $\text{val}(\text{DL}_d)$. 然而, 如下例所示, 这可能是不对的, 上述问题的答案是否定的.

例 4.23 令 $X = \mathbb{R}^2$, $U = \mathbb{R}^2$, 考虑问题: 极小化 $f(x) = -x_2$, 受约束于 $x_2 + x_1^2 \leqslant u_1$ 与 $x_2 - x_2^2 \leqslant u_2$. 对 $u_0 = (0,0)$, 此规划有唯一的最优解 $x_0 = (0,0)$ 满足二阶增长条件. Mangasarian-Fromovitz 约束规范在这一点处成立, (有界的)Lagrange 乘子集为

$$\Lambda(x_0, u_0) = \{(\lambda_1, \lambda_2) : \lambda_1 + \lambda_2 = 1,\ \lambda_1 \geqslant 0,\ \lambda_2 \geqslant 0\}.$$

考虑方向 $d = (1, 0)$. 则对 $u = td$, $t \geqslant 0$, 这一规划有两个最优解 $((t/2)^{1/2}, t/2)$ 与 $(-(t/2)^{1/2}, t/2)$, 最优值函数 $v(td) = -t/2$, 因此 $v'(u_0, d) = -\frac{1}{2}$. 另一方面,

$$\text{val}(\text{PL}_d) = \text{val}(\text{DL}_d) = \min_{\lambda \in \Lambda(x_0, u_0)} d \cdot \nabla_u L(x_0, \lambda, u_0) = 0.$$

后面将会看到, $v(u)$ 的方向导数的下界基于二阶分析. 然而, 若非扰动问题 (P_{u_0}) 是凸的, 则依然可以用一阶分析的方法得到 $v(u)$ 的方向导数. 此种情况, $\Lambda(x_0, u_0)$ 对所有的 $x_0 \in \mathcal{S}(u_0)$ 是相同的, 它是 (P_{u_0}) 的对偶问题的最优解集. 所以, 可省掉 x_0, 将此集合记为 $\Lambda(u_0)$. 下述定理是 Gol'shtein 的一个定理在无穷维情况及方向正则性情形下的推广. 在各种问题中, 这一定理中的假设 (iii) 是事前可以验证的.

定理 4.24 设

(i) 问题 (P_{u_0}) 是凸的, 具有非空的最优的解集 $\mathcal{S}(u_0)$.

(ii) 对所有的 $x_0 \in \mathcal{S}(u_0)$, 沿方向 d, 方向正则性条件成立.

(iii) 若对某一序列 $t_n \downarrow 0$, $u_n := u_0 + t_n d + o(t_n)$, 则 (P_{u_n}) 具有 $o(t_n)$ 最优解 x_n, 满足序列 $\{x_n\}$(在强拓扑下) 具有极限点 $x_0 \in \mathcal{S}(u_0)$.

则最优值函数 $v(\cdot)$ 在 u_0 处沿方向 d 是 Hadamard 方向可微的, 且

$$v'(u_0, d) = \inf_{x \in \mathcal{S}(u_0)} \sup_{\lambda \in \Lambda(u_0)} D_u L(x, \lambda, u_0) d. \tag{4.52}$$

进一步, 若对 $t_n \downarrow 0$, $d_n \to d, x_n$ 是 $(P_{u_0 + t_n d_n})$ 的 $o(t_n)$ 最优解, 且 $x_n \to x_0 \in \mathcal{S}(u_0)$, 则 $x_0 \in \mathcal{S}_1(u_0, d)$, 其中

$$\mathcal{S}_1(u_0, d) := \operatorname*{argmin}_{x \in \mathcal{S}(u_0)} \left\{ \phi(x) := \sup_{\lambda \in \Lambda(u_0)} D_u L(x, \lambda, u_0) d \right\}. \tag{4.53}$$

证明　由于 $\mathrm{val}(\mathrm{PL}_d) = \mathrm{val}(\mathrm{DL}_d)$, 且 (4.51) 对任意的 $x \in \mathcal{S}(u_0)$ 成立, 可得

$$v'_+(u_0, d) \leqslant \inf_{x\in\mathcal{S}(u_0)} \sup_{\lambda\in\Lambda(u_0)} D_u L(x, \lambda, u_0)d. \tag{4.54}$$

尤其, 若 $\Lambda(u_0)$ 是空集, 则 $v'_+(u_0, d) = -\infty$, (4.52) 显然成立. 所以, 可设 $\Lambda(u_0) \neq \varnothing$. 为推导最优值函数的下方向估计, 现在用 (P_{u_0}) 的凸性.

考虑序列 $t_n \downarrow 0$, 令 $u_n := u_0 + t_n d + o(t_n)$, $\{x_n\}$ 是相应的 (P_{u_n}) 的 $o(t_n)$ 最优解, 它有极限点 $x_0 \in \mathcal{S}(u_0)$. 若有必要, 可抽取一子序列, 不妨设 $x_n \to x_0$. 令 $\lambda \in \Lambda(u_0)$. 由于 $\Lambda(u_0) \subset N_K(G(x_0, u_0))$, 有 $\langle \lambda, G(x_n, u_n) - G(x_0, u_0)\rangle \leqslant 0$, 则

$$f(x_n, u_n) - f(x_0, u_0) \geqslant L(x_n, \lambda, u_n) - L(x_0, \lambda, u_0). \tag{4.55}$$

由 (P_{u_0}) 的凸性, 一阶最优性条件 (3.8) 可推出 $x_0 \in \operatorname*{argmin}_{x\in X} L(x, \lambda, u_0)$, 因此 $L(x_n, \lambda, u_0) \geqslant L(x_0, \lambda, u_0)$. 结果有

$$f(x_n, u_n) - f(x_0, u_0) \geqslant L(x_n, \lambda, u_n) - L(x_n, \lambda, u_0). \tag{4.56}$$

因为 $f(x_n, u_n) = v(u_n) + o(t_n)$, 有 $f(x_0, u_0) = v(u_0)$, 由中值定理及 $L(x, \lambda, u)$ 的连续性, 由 (4.56) 可得, 存在 $u'_n \in [u_0, u_n]$,

$$v(u_n) - v(u_0) \geqslant t_n D_u L(x_n, \lambda, u'_n)d = t_n D_u L(x_0, \lambda, u_0)d + o(t_n),$$

因此

$$\liminf_{n\to\infty} \frac{v(u_n) - v(u_0)}{t_n} \geqslant D_u L(x_0, \lambda, u_0)d.$$

由于 $\lambda \in \Lambda(u_0)$ 是任意的, 得到

$$\liminf_{n\to\infty} \frac{v(u_n) - v(u_0)}{t_n} \geqslant \sup_{\lambda\in\Lambda(u_0)} D_u L(x_0, \lambda, u_0)d. \tag{4.57}$$

同 (4.54) 相联系, 这可推出

$$\lim_{n\to\infty} \frac{v(u_n) - v(u_0)}{t_n} = \sup_{\lambda\in\Lambda(u_0)} D_u L(x_0, \lambda, u_0)d.$$

当然得到 $x_0 \in \mathcal{S}_1(u_0, d)$, 公式 (4.52) 成立, 这完成了证明. □

现在讨论具有下述形式

$$v(u(t)) - v(u_0) \geqslant t\ \mathrm{val}(\mathrm{PL}_d) + o(t) \tag{4.58}$$

的下界 (其中 $u(t) := u_0 + td + o(t)$, $t \geqslant 0$) 与近似解的稳定性之间的关系. 令 $x_0 \in \mathcal{S}(u_0)$ 满足方向正则性条件成立且相应的线性化问题 (PL_d) 的最优值是有限的. 令 h

是 (PL_d) 的 ε 最优解. 则由定理 4.9, 存在可行的路径 $x(t) = x_0 + th + o(t) \in \Phi(u(t))$, 有

$$v(u(t)) - v(u_0) \leqslant f(x(t), u(t)) - f(x_0, u_0) \leqslant t \, \mathrm{val}(\mathrm{PL}_d) + t\varepsilon + o(t).$$

因此, 若还有 (4.58) 的下确界成立, 则

$$v(u(t)) \geqslant f(x(t), u(t)) - t\varepsilon + o(t).$$

有对任何 $\varepsilon' > \varepsilon$, 对 $t \geqslant 0$ 充分小, $x(t)$ 是 $(P_{u(t)})$ 的 $(t\varepsilon')$ 最优解. 显然, $\|x(t) - x_0\| = O(t)$.

所以, 下界 (4.58) 可推出等式 $v'(u_0, d) = \mathrm{val}(\mathrm{DL}_d)$ 且对任何 $\varepsilon > 0$, 存在 $(P_{u(t)})$ 的 $(t\varepsilon)$ 最优解 $x(t)$ 满足 $\|x(t) - x_0\| = O(t)$. 可以证明, 在某一意义下, 相反的结论也是成立的, 即由在 x_0 处 Lipschitz 稳定的 $(P_{u(t)})$ 的近似最优解的存在性可推出 $v'(u_0, d) = \mathrm{val}(\mathrm{DL}_d)$.

定理 4.25 设

(i) 问题 (P_{u_0}) 具有非空的最优解集 $\mathcal{S}(u_0)$.

(ii) 对所有的 $x_0 \in \mathcal{S}(u_0)$, 沿方向 d 的方向正则性条件成立.

(iii) 对任何 $u(t) = u_0 + td + o(t)$, $t > 0$ 充分小, 问题 $(P_{u(t)})$ 具有 $o(t)$ 最优解 $\bar{x}(t)$ 满足 $\mathrm{dist}(\bar{x}(t), \mathcal{S}(u_0)) = O(t)$.

(iv) 对任何 $t_n \downarrow 0$, 序列 $\{\bar{x}(t_n)\}$ 有极限点 $x_0 \in \mathcal{S}(u_0)$.

则最优值函数在 u_0 处沿方向 d 是 Hadamard 方向可微的, 且

$$v'(u_0, d) = \inf_{x \in \mathcal{S}(u_0)} \sup_{\lambda \in \Lambda(x_0, u_0)} D_u L(x, \lambda, u_0) d. \tag{4.59}$$

证明 我们像证明定理 4.24 那样进行证明. 上界结果 (4.54) 的成立是不涉及凸性假设的, 因此在目前情况下也是成立的. 若对某一 $x \in \mathcal{S}(u_0)$, 有 $\Lambda(x, u_0)$ 是空集, 则由上界 (4.54), 结论为真. 因此, 设对所有的 $x \in \mathcal{S}(u_0)$ 有 $\Lambda(x, u_0)$ 是非空的.

为证明下界结果, 我们如下进行. 令 $u_n := u_0 + t_n d + o(t_n)$, $x_n := \bar{x}(t_n)$ 收敛到某一 $x_0 \in \mathcal{S}(u_0)$. 因为由假设有 $\mathrm{dist}(x_n, \mathcal{S}(u_0)) = O(t_n)$, 则存在 $x'_n \in \mathcal{S}(u_0)$ 满足 $\|x_n - x'_n\| = O(t_n)$. 令 $\lambda'_n \in \Lambda(x'_n, u_0)$ 是对应点 x'_n 的对偶问题 (DL_d) 的最优解, 即

$$\lambda'_n \in \underset{\lambda \in \Lambda(x'_n, u_0)}{\mathrm{argmax}} \; D_u L(x'_n, \lambda, u_0) d.$$

因为方向正则性条件在 x'_n 处成立, 由命题 4.21 得这样的最优解存在. 进一步, 由于在点 $x \in \mathcal{S}(u_0)$ 处的方向正则性条件可推出相应的线性化问题 (PL_d) 的 Robinson 约束规范, 又因为对任何 $x \in \mathcal{S}(u_0)$, $D_u L(x, \cdot, u_0) d$ 在 $\Lambda(x, u_0)$ 的极大值点集合

与 (PL_d) 的 Lagrange 乘子集合重合, 于是得到这些集合[①]在 x_0 的邻域内所有的 $x \in \mathcal{S}(u_0)$ 是一致有界的. 后面将证明这个结果 (见 4.4.3 节中的命题 4.43). 得到 λ_n' 也是一致有界的. 则有 (见 (4.55))

$$f(x_n, u_n) - f(x_n', u_0) \geqslant L(x_n, \lambda_n', u_n) - L(x_n', \lambda_n', u_0). \tag{4.60}$$

因为 λ_n' 是一致有界的, 有

$$L(x_n, \lambda_n', u_n) - L(x_n, \lambda_n', u_0) = t_n D_u L(x_n', \lambda_n', u_0) + o(t_n), \tag{4.61}$$

由于 $\|x_n - x_n'\| = O(t_n)$,

$$L(x_n, \lambda_n', u_0) - L(x_n', \lambda_n', u_0) = D_x L(x_n', \lambda_n', u_0)(x_n - x_n') + o(t_n).$$

由一阶必要性条件 $D_x L(x_n', \lambda_n', u_0) = 0$, 由于 $v(u_n) = f(x_n, u_n) + o(t_n)$ 及 $v(u_0) = f(x_n', u_0)$, 由 (4.60) 与 (4.61) 得

$$v(u_n) - v(u_0) \geqslant t_n D_u L(x_n', \lambda_n', u_0) + o(t_n),$$

这证得下界结果. □

考虑下述 (一阶) 结果, 它对 (不必是凸的) 一般规划是成立的. 注意, 下述定理假设 (ii) 等同于定理 4.24 假设 (iii), 下 Hadamard 方向导数与下方向上图导数是一样的 (见 2.2.3 节).

定理 4.26　设

(i) *在每一点 $x \in \mathcal{S}(u_0)$ 处的 Robinson 约束规范成立.*

(ii) *若 $u_n := u_0 + t_n d + o(t_n)$, $t_n \downarrow 0$, 则 (P_{u_n}) 具有 $o(t_n)$ 最优解 x_n, 具有强极限点 $x_0 \in \mathcal{S}(u_0)$.*

则对每一 $d \in U$,

$$\inf_{x \in \mathcal{S}(u_0)} \inf_{\lambda \in \Lambda(x, u_0)} D_u L(x, \lambda, u_0) d \leqslant v_-'(u_0, d) \leqslant v_+'(u_0, d) \leqslant \inf_{x \in \mathcal{S}(u_0)} \sup_{\lambda \in \Lambda(x, u_0)} D_u L(x, \lambda, u_0) d, \tag{4.62}$$

其中下导数 $v_-'(u_0, d)$ 与上导数 $v_+'(u_0, d)$ 理解为 Hadamard 意义下的方向导数.

若还有对每一 $x \in \mathcal{S}(u_0)$, $\Lambda(x, u_0) = \{\overline{\lambda}(x)\}$ 是单点集, 则 $v(u)$ 在 u_0 处是 Hadamard 方向可微的, 且

$$v'(u_0, d) = \inf_{x \in \mathcal{S}(u_0)} D_u L(x, \overline{\lambda}(x), u_0) d, \tag{4.63}$$

① 即极大值点集合.

进一步, 若对某一 $t_n \downarrow 0$, $d_n \to d$, x_n 是 $(P_{u_0+t_nd_n})$ 的 $o(t_n)$ 最优解, 且 $x_n \to x_0 \in \mathcal{S}(u_0)$, 则 $x_0 \in \mathcal{S}_1(u_0,d)$, 其中

$$\mathcal{S}_1(u_0,d) := \operatorname*{argmin}_{x\in\mathcal{S}(u_0)} D_uL(x,\overline{\lambda}(x),u_0)d. \tag{4.64}$$

证明 由假设 (ii), 有 $\mathcal{S}(u_0) \neq \varnothing$. 因为 (4.62) 中的第二个不等式是显然的, 第三个不等式由定理 4.24 得到, 只需要证第一个不等式.

考虑由假设 (ii) 给定的 $t_n \downarrow 0$, 设 $u_n := u_0 + t_nd + o(t_n)$, $\{x_n\}$ 是相应的 $o(t_n)$ 最优解序列, 收敛到 $x_0 \in \mathcal{S}(u_0)$. 则

$$f(x_n,u_n) \leqslant v(u_n) + o(t_n). \tag{4.65}$$

因为在 x_0 处的 Robinson 约束规范成立, 有 (由命题 2.97)

$$D_xG(x_0,u_0)X - T_K(G(x_0,u_0)) = Y,$$

则存在 $h \in X$ 满足

$$DG(x_0,u_0)(h,-d) \in T_K(G(x_0,u_0)). \tag{4.66}$$

这意味着存在映射 $r:\mathbb{R}_+ \to Y$ 满足 $r(\tau) = o(\tau)$, $\tau \geqslant 0$,

$$G(x_0,u_0) + \tau DG(x_0,u_0)(h,-d) + r(\tau) \in K. \tag{4.67}$$

置

$$Z(x,u,\tau) := \tau^{-1}[G(x_0,u_0) - G(x,u)] + DG(x_0,u_0)(h,-d) + \tau^{-1}r(\tau). \tag{4.68}$$

由 (4.67), 对所有的 $(x,u) \in X \times U$,

$$G(x,u) + \tau Z(x,u,\tau) \in K.$$

因为 $G(x_n,u_n) \in K$, 由 K 的凸性有

$$G(x_n,u_n) + tZ(x_n,u_n,\tau) \in K, \quad \forall t \in [0,\tau]. \tag{4.69}$$

固定 $\varepsilon > 0$, 令 $\tau \geqslant 0$ 充分小, 使得 $\|\tau^{-1}r(\tau)\| \leqslant \varepsilon/4$. 由于 $(x_n,u_n) \to (x_0,u_0)$, 对充分大的 n, 由 (4.68) 得

$$\|DG(x_n,u_n)(h,-d) - Z(x_n,u_n,\tau)\| \leqslant \frac{1}{2}\varepsilon.$$

对充分大的 n, $t_n \leqslant \tau$, 由 (4.69) 有

$$\begin{aligned} G(x_n + t_n h, u_0) &= G(x_n, u_n) + t_n DG(x_n, u_n)(h, -d) + o(t_n) \\ &\in G(x_n, u_n) + t_n Z(x_n, u_n, \tau) + t_n \varepsilon B_Y \\ &\in K + t_n \varepsilon B_Y. \end{aligned}$$

由于 Robinson 约束规范在 (x_0, u_0) 处成立, 由定理 2.87 得到, 存在 $\hat{x}_n \in \Phi(u_0)$ 满足 $x_n + t_n h - \hat{x}_n = O(t_n \varepsilon)$. 结果, 由于 $v(u_n) = f(x_n, u_n) + o(t_n)$, 又由 $Df(x, u)$ 的连续性,

$$\begin{aligned} v(u_0) \leqslant f(\hat{x}_n, u_0) &= f(x_n + t_n h, u_0) + O(t_n \varepsilon) \\ &= f(x_n + t_n h, u_n - t_n d) + O(t_n \varepsilon) \\ &= v(u_n) + t_n Df(x_0, u_0)(h, -d) + O(t_n \varepsilon). \end{aligned}$$

因为 h 是满足 (4.66) 的任意方向, 在 (PL_{-d}) 与 (DL_{-d}) 间不存在对偶间隙, 有

$$v(u_0) - v(u_n) \leqslant t_n \sup_{\lambda \in \Lambda(x_0, u_0)} D_u L(x_0, \lambda, u_0)(-d) + O(t_n \varepsilon) \tag{4.70}$$

对任意 $\varepsilon > 0$ 成立, 得到

$$\liminf_{n \to \infty} \frac{v(u_n) - v(u_0)}{t_n} \geqslant \inf_{\lambda \in \Lambda(x_0, u_0)} D_u L(x_0, \lambda, u_0) d, \tag{4.71}$$

于是 (4.62) 中的第一个不等式得证.

在 $\Lambda(x, u_0) = \{\overline{\lambda}(x)\}$ 是单点集的情况, (4.63) 显然可由 (4.62) 得到, 最后一个结论由 (4.71) 得到. □

后面 (见 4.4.3 节) 将讨论确保 Lagrange 乘子唯一性的条件.

让我们做如下注记来结束这一节, 注意到, 若问题 (P_u) 关于 x 与 u 是凸的, 即具有 2.5 节中讨论的形式, 则 $v(u)$ 的方向可微性由定理 2.142 与 2.151 来描述. 尤其, 考虑问题

$$(\mathrm{P}_u) \qquad \min_{x \in X} f(x) \quad \text{s. t.} \quad G(x) + u \in K, \tag{4.72}$$

以 $u \in Y$ 为参数. 对 $u = 0$, (4.72) 即为 (2.291) 定义的问题 (P). (P) 的对偶问题 (D) 由 (2.298) 定义. 下述结果是定理 2.151 与命题 2.152 的一个结论.

命题 4.27　*设 $u = 0$ 时的问题 (P) 是凸的, val(P) 是有限的, 正则性条件 (2.312) 成立. 则问题 (4.72) 的最优值函数 $v(u) = \mathrm{val}(\mathrm{P}_u)$ 在 $u = 0$ 处是 Hadamard 方向可微的, 且*

$$v'(0, d) = \sup_{\lambda \in \mathcal{S}(D)} \langle \lambda, d \rangle, \tag{4.73}$$

其中 $\mathcal{S}$(D) 是对偶问题 (D) 的最优解集.

在上述命题的假设之下, (P) 具有最优解 x_0, 对偶问题 (D) 的最优解集与 (P) 的 Lagrange 乘子集是重合的, 即 $\mathcal{S}(\mathrm{D})=\Lambda(x_0)$ (见定理 3.4). 若 $u=\Psi(z)$, 其中 $\Psi: Z\to U$ 是方向可微的, 则相应的最优值函数 $\bar{v}(z)=v(\Psi(z))$ 的方向导数可以由链式法则 (2.39) 通过 (4.73) 来计算.

例 4.28 考虑问题

$$\min_{x\in\mathbb{R}^2} x_1 \quad \text{s. t.} \quad -x_1+\mathrm{e}^{x_2}+u\leqslant 0, \tag{4.74}$$

以 $u\in\mathbb{R}$ 为参数 (与例 2.154 比较). 对 $u=0$, 这是正常的凸问题, 其对偶只有一可行点 $\lambda=1$. Slater 条件成立, 原始问题与对偶问题间不存在对偶间隙. 容易验证 $v(u)=u$, 从而公式 (4.73) 成立, 尽管原始问题 (4.74) 不具有最优解.

例 4.29 考虑例 3.20 中的情况, 其中 $X=\mathbb{R}$, $Y=L_2[0,1]$, $f(x)=x$, $G(x)(t)=t-x$. 这个例子中 $x_0=1$ 是 (P) 的最优解, 且相应的广义的 Lagrange 乘子集合是空集. 这一问题的 Lagrange 函数是

$$L(x,\lambda)=x+\int_0^1(t-x)\lambda(t)\mathrm{d}t,$$

极锥K^- 由几乎处处非负值函数 $\lambda\in L_2[0,1]$ 组成. 则对 $\lambda\in K^-$, 有

$$\inf_{x\in\mathbb{R}} L(x,\lambda)=\begin{cases}\displaystyle\int_0^1 t\lambda(t)\mathrm{d}t, & \text{若}\displaystyle\int_0^1\lambda(t)\mathrm{d}t=1,\\ -\infty, & \text{否则}.\end{cases}$$

则对偶问题可以写为下述形式

$$\max_{\lambda\in L_2[0,1]}\int_0^1 t\lambda(t)\mathrm{d}t \quad \text{s. t.} \quad \lambda\in K^-\text{且}\int_0^1\lambda(t)\mathrm{d}t=1. \tag{4.75}$$

对偶问题 (4.75) 的最优值是 1, 因此这里的原始问题与对偶问题间没有对偶间隙. 进一步, 对偶问题 (4.75) 的最优解集是空集, 因此 (见定理 2.142(ii)) 相应的最优值函数 $v(u)$ 在 $u=0$ 处不是次可微的.

注意, 扰动问题的可行集 $\Phi(u)$ 由对几乎每一 $t\in[0,1]$ 成立 $t+u(t)\leqslant x$ 的 $x\in\mathbb{R}$ 构成. 结果, 对所有的 $u\in L_2[0,1]$, $v(u)>-\infty$, 若扰动函数 $u\in L_2[0,1]$ 在 $[0,1]$ 上是本质上方无界的, 则 $\Phi(u)$ 是空集, 从而 $v(u)=+\infty$. 因此, 此例中, 最优值函数在 Y 的稠密子集上是有限值的, 不是连续的, 因而在 $u=0$ 处不是次可微的. 进一步, 考虑函数 $d(t):=-(1-t)^{-1/3}\in L_2[0,1]$. 对 $\tau>0$, 可行集 $\Phi(\tau d)$ 由满足 $x\geqslant\bar{x}(\tau)$ 的 $x\in\mathbb{R}$ 组成, 其中 $\bar{x}(\tau)=\max\limits_{t\in[0,1]}\{t+\tau d(t)\}$, 且 $v(\tau d)=\bar{x}(\tau)$. 直接计算可得, 存在 $\kappa>0$, $\bar{x}(\tau)=1-\kappa\tau^{3/4}+o(\tau)$, 因而 $v'(0,d)=-\infty$. 这给出了最优值函数 $v(u)$ 在 $u=0$ 处不是次可微的另一解释.

4.4 最优解与 Lagrange 乘子的量化稳定性

这一节设 X,Y 与 U 是 Banach 空间, 讨论问题 (P_u) 的最优解与 Lagrange 乘子的连续性质. 我们尤其对最优解多值函数 $\mathcal{S}: U \to 2^X$ 的 Hölder 与 Lipschitz 性质感兴趣.

定义 4.30 称 $\mathcal{S}(u)$ 在点 $u_0 \in U$ 处是 α 度 $(\alpha > 0)$Hölder 稳定的, 若存在常数 $c>0$ 满足对 u_0 的邻域中所有的 u, 有

$$\mathcal{S}(u) \subset \mathcal{S}(u_0) + c\|u-u_0\|^\alpha B_X. \tag{4.76}$$

若 $\mathcal{S}(u)$ 是度 $\alpha=1$ 的 Hölder 稳定的, 即 $\mathcal{S}(u)$ 在 u_0 处是上 Lipschitz 的, 称 $\mathcal{S}(u)$ 是 Lipschitz 稳定的.

下面讨论 (P_u) 的 ε 最优解的 Hölder 与 Lipschitz 稳定性.

例 4.31 考虑 $f(x,u) := x^{2m} - ux$, 其中 $u \in \mathbb{R}$, m 是正整数, 令 $\Phi(u) := [-1,1]$, $\forall u \in \mathbb{R}$. 则对任何 $u \in (-1,1)$, $\mathcal{S}(u) = \{u^{1/(2m-1)}\}$ 是单点集. 因此, 若 $m \geqslant 2$, $\mathcal{S}(u)$ 在 $u=0$ 处不是 Lipschitz 稳定的.

上述例子表明, 为建立最优解的 Lipschitz 性质, 除了所涉及函数的连续性与一阶可微性性质, 还需要额外的一些假设. 由此可以看到, $\mathcal{S}(u)$ 的量化稳定性与 (P_u) 的二阶分析是密切相关的.

4.4.1 固定可行集情况的 Lipschitz 稳定性

当可行集 $\Phi(u) \equiv \Phi$ 是非扰动的, 即与 u 无关时, 可以相对容易地给出 $\mathcal{S}(u)$ 的 Lipschitz 稳定性的一般性充分性条件. 考虑两个最优化问题

$$\min_{x\in\Phi} f(x) \tag{4.77}$$

与

$$\min_{x\in\Phi} g(x), \tag{4.78}$$

其中 f, $g: X \to \mathbb{R}$. 设问题 (4.77) 具有非空的最优解集 S_0. 视 (4.78) 中的函数 g 是函数 f 的扰动, 我们感兴趣的是导出 (4.78) 的 ε 最优解 $\bar{x}$ 与集合 S_0 的距离的上界. 在 S_0 处的二阶增长条件成立指存在 S_0 的一个邻域 N 及常数 $c>0$, 满足

$$f(x) \geqslant f_0 + c[\mathrm{dist}(x,S_0)]^2, \quad \forall x \in \Phi \cap N, \tag{4.79}$$

其中 $f_0 := f(S_0) = \inf_{x\in\Phi} f(x)$. 一阶 ($\gamma$- 阶) 增长性条件可类似定义 (见定义 3.1).

命题 4.32 设

(i) 二阶增长条件(4.79) 成立.

(ii) 在 $\Phi \cap N$ 上, 差函数 $g(\cdot) - f(\cdot)$ 是模为 κ 的 Lipschitz 连续的.

令 $\bar{x} \in N$ 是问题 (4.78) 的 ε 解. 则

$$\operatorname{dist}(\bar{x}, S_0) \leqslant c^{-1}\kappa + c^{-1/2}\varepsilon^{1/2}. \tag{4.80}$$

证明 考虑差函数 $h := g - f$, 一点 $x_0 \in S_0$. 则由于 $\bar{x}$ 是 (4.78) 的 ε 解, 有 $g(\bar{x}) \leqslant g(x_0) + \varepsilon$, 因此

$$\begin{aligned} f(\bar{x}) - f(x_0) &= h(x_0) - h(\bar{x}) + g(\bar{x}) - g(x_0) \\ &\leqslant h(x_0) - h(\bar{x}) + \varepsilon \leqslant \kappa\|x_0 - \bar{x}\| + \varepsilon. \end{aligned} \tag{4.81}$$

进一步, 对任何 $\eta > 0$, 可选取 $x_0 \in S_0$ 满足

$$\|\bar{x} - x_0\| \leqslant \operatorname{dist}(\bar{x}, S_0) + \eta.$$

所以, 由二阶增长条件 (4.79), 得

$$f(\bar{x}) - f(x_0) \geqslant c(\|\bar{x} - x_0\| - \eta)^2.$$

结合 (4.81) 可推出

$$c(\|\bar{x} - x_0\| - \eta)^2 \leqslant \kappa\|\bar{x} - x_0\| + \varepsilon.$$

通过对上面的不等式取极限可得

$$c \operatorname{dist}(\bar{x}, S_0)^2 \leqslant \kappa \operatorname{dist}(\bar{x}, S_0) + \varepsilon. \tag{4.82}$$

通过解上述不等式, 得

$$\operatorname{dist}(\bar{x}, S_0) \leqslant \frac{1}{2}c^{-1}\kappa + \sqrt{\left(\frac{1}{2}c^{-1}\kappa\right)^2 + c^{-1}\varepsilon} \leqslant c^{-1}\kappa + c^{-1/2}\varepsilon^{1/2},$$

这证得结论. □

关于解的扰动解的上界, 结果 (4.80) 是非常适宜的, 因为它几乎不需要所考虑的最优化问题的任何结构.

例 4.33 设 X 是 Hilbert 空间, $f : X \to \overline{\mathbb{R}}$ 是正常的 l.s.c. 凸函数, 考虑 f 的 Moreau-Yosida 正则化, 即

$$\hat{f}_\varepsilon(u) := \min_{x \in X}\left\{f(x) + \frac{1}{2}\varepsilon\|x - u\|^2\right\}, \quad \varepsilon > 0. \tag{4.83}$$

注意到 $\hat{f}_\varepsilon(u)$ 可以写为下述的形式

$$\hat{f}_\varepsilon(u) = \min_{x\in C}\Big\{ f(x) + \frac{1}{2}\varepsilon\|x-u\|^2 \Big\}, \tag{4.84}$$

其中 $C := \mathrm{dom} f$. 尤其, 若 f 是一指示函数$f(\cdot) = I_C(\cdot)$, 则 $\hat{f}_\varepsilon(u) = \frac{1}{2}\varepsilon\,\mathrm{dist}(u,C)^2$.

考虑函数

$$F_u(x) := f(x) + \frac{1}{2}\varepsilon\|x-u\|^2.$$

有 $F_u(\cdot)$ 是正常的下半连续的凸函数, 因此它是弱 * 下半连续的. 进一步, 其水平集是有界的. 事实上, 因为 f 是正常的 l.s.c. 凸函数, 其共轭函数 f^* 也是正常的 (见命题 2.112), 因此, 存在点 $x^* \in X^*$ 满足 $\alpha := f^*(x^*)$ 是有限的, 有 $f(\cdot) \geqslant \langle x^*, \cdot\rangle - \alpha$, 因此

$$F_u(x) \geqslant \langle x^*, x\rangle + \frac{1}{2}\varepsilon\|x-u\|^2 - \alpha. \tag{4.85}$$

因为 (4.85) 右端的二次函数的水平集是有界的, 得 $F_u(\cdot)$ 的水平集也是有界的. 因为 Hilbert 空间 X 是自反的, 所以上述问题 (4.83) 总具有最优解 $\bar{x}(u)$(见推论 2.29). 进一步, 函数 $F_u(\cdot)$ 是严格凸的, 其最优解 $\bar{x}(u)$ 是唯一的.

这里还有函数 $F_u(\cdot)$ 是凸函数与强凸函数的和, 其 Hesse 阵在每一点处均是一致正定的, 因此在整个空间 X 上于 $\bar{x}(u)$ 处的二阶增长条件成立, 其常数是 $c = \varepsilon/2$. 进一步, 差函数 $F_{u_1}(\cdot) - F_{u_2}(\cdot)$ 在 X 上是模为 $\varepsilon\|u_1 - u_2\|$ 的 Lipschitz 连续函数. 由 (4.80) 得

$$\|\bar{x}(u_1) - \bar{x}(u_2)\| \leqslant 2\|u_1 - u_2\|, \tag{4.86}$$

即 $\bar{x}(\cdot)$ 在 X 上是 Lipschitz 连续的. 进一步, 由 Danskin 定理 4.13,

$$D\hat{f}_\varepsilon(u) = D_u\left(f(\bar{x}) + \frac{1}{2}\varepsilon\|\bar{x}-u\|^2 \right) = \varepsilon(u - \bar{x}), \tag{4.87}$$

其中 $\bar{x} = \bar{x}(u)$ (即使集合 $C := \mathrm{dom} f$ 不是紧致的, 因为相应的极小点总是存在的, 因而是连续的, 所以 Danskin 定理对这种情况亦是可以应用的). 由 (4.86), $\bar{x}(u)$ 是 Lipschitz 连续的, 得 $\hat{f}_\varepsilon(u)$ 在 X 上具有 Lipschitz 连续的导数, 因而是 Fréchet 可微的.

例 4.34 设 X 是 Banach 空间, $f: X \to \overline{\mathbb{R}}$ 是正常的 l.s.c. 凸函数. 回顾共轭函数 $f^*(x^*)$ 的定义 (2.210) 及次微分可以写为下述形式 (见 (2.232))

$$\partial f(x) = \underset{x^*\in X^*}{\mathrm{argmin}}\Big\{ f^*(x^*) - \langle x^*, x\rangle \Big\}. \tag{4.88}$$

为了建立其 Lipschitz 连续性性质, 我们可以用上述 $\partial f(x)$ 的变分表示. 令 $x_0 \in X$ 为给定点, 满足 $\partial f(x_0)$ 是非空的, 设下述二阶增长条件成立：存在 $\partial f(x_0)$ 的邻域

N 及常数 $c>0$ 满足

$$f^*(x^*)-\langle x^*,x_0\rangle \geqslant f_0+c[\mathrm{dist}(x^*,\partial f(x_0))]^2,\quad \forall x^*\in N, \tag{4.89}$$

其中, $f_0:=\inf\limits_{x^*\in X^*}\left\{f^*(x^*)-\langle x^*,x_0\rangle\right\}$. 由命题 4.32 中的上界 (4.80) 可得, 若条件 (4.89) 成立, 则只要 $\partial f(x)\subset N$, 就有

$$\partial f(x)\subset \partial f(x_0)+c^{-1}\|x-x_0\|B_{X^*}. \tag{4.90}$$

尤其, 设 X 是 Hilbert 空间, 因而 X^* 等同于 X 本身, 设共轭函数 f^* 是强凸的 (见定义 2.32). 则对任意 $x,\ x_0\in X$ 及 $z\in\partial f^*(x_0)$ 有

$$f^*(x)-f^*(x_0)\geqslant \langle z,x-x_0\rangle+\frac{\alpha}{2}\|x-x_0\|^2. \tag{4.91}$$

由上述二阶增长条件可推出, 在 (4.88) 右端中括号里的函数的水平集是有界的, 因此, 对所有的 $x\in X$, 有 $\partial f(x)$ 是非空的, 进一步, 最小值在唯一点处达到, 从而 $\partial f(x)=\{\nabla f(x)\}$ 是单点集. 由上述讨论, 得到 $\nabla f(\cdot)$ 在 X 上是模 $2\alpha^{-1}$ 的 Lipschitz 连续的.

注 4.35 若 γ 阶增长条件对于问题 (4.77) 于 S_0 成立, 则由与命题 4.32 的证明完全相同的推证可得到与不等式 (4.82) 相类似的如下结果

$$c\ [\mathrm{dist}(\bar{x},S_0)]^\gamma\leqslant \kappa\ \mathrm{dist}(\bar{x},S_0)+\varepsilon. \tag{4.92}$$

尤其, 若 $\gamma=1$, 即一阶增长条件成立, 若 (差函数 $g-f$ 的)Lipschitz 常数小于 c, 则由 (4.92) 可得

$$\mathrm{dist}(\bar{x},S_0)\leqslant (c-\kappa)^{-1}\varepsilon. \tag{4.93}$$

尤其, 对 $\varepsilon=0$, 扰动问题的局部最优解属于集合 S_0.

考虑固定可行集 $\Phi(u)\equiv\Phi$ 的问题 (P_u)(由 (4.26) 给出). 将非扰动问题 (P_{u_0}) 视为问题 (4.77), 把问题 (P_u)(其中 u 是给定的) 视为问题 (4.78). 即令 $f(\cdot)=f(\cdot,u_0)$, $g(\cdot)=f(\cdot,u)$. 若函数 $f(x,u)$ 是关于 x 的连续可微的, 则 (由中值定理) 差函数 $f(\cdot,u)-f(\cdot,u_0)$ 在 N 上的 Lipschitz 常数 $\kappa(u)$ 可估计为

$$\kappa(u)\leqslant \sup_{x\in N}\|D_xf(x,u)-D_xf(x,u_0)\|.$$

所以, 下述结果是命题 4.32 的一个直接的结论.

命题 4.36 设 $\Phi(u)\equiv\Phi$, (对非扰动问题的) 二阶增长条件 (4.79) 在 $S_0:=\mathcal{S}(u_0)$ 处成立, $f(x,u)$ 是 x 的连续可微函数, $D_xf(x,\cdot)$ 是 u_0 的邻域上 Lipschitz 连续的, 且相应的 Lipschitz 常数与 $x\in N$ 无关, 令 $t_n\downarrow 0$, $\{x_n\}$ 满足 $x_n\in\Phi\cap N$

且 $f(x_n, u_n) \leqslant v(u_n) + O(\|u-u_0\|^2)$ 对某一序列 $u_n \to u_0$ 成立. 则 $\operatorname{dist}(x_n, S_0) = O(\|u_n - u_0\|)$, 即在邻域 N 中的任何 $O(\|u-u_0\|^2)$ 最优解路径在 u_0 处是 Lipschitz 稳定的①.

在某些特殊的情况, 可以把可行集 $\Phi(u)$ 依赖于 u 的问题转化为可行集不依赖于 u 的问题. 在 4.6 节将讨论这些情况.

4.4.2 抽象约束下的 Hölder 稳定性

现在考虑可行集 $\Phi(u)$ 也依赖于 u 的一般的情况. 在例 4.23 中, Mangasarian-Fromovitz 约束规范及二阶充分性条件 (对非扰动问题)

$$\sup_{\lambda\in\Lambda(x_0)} D_{xx}^2 L(x_0,\lambda)(h,h) > 0, \quad \forall h \in C(x_0)\setminus\{0\} \tag{4.94}$$

成立, 然而 (两个) 最优解在 $u=0$ 处不是 Lipschitz 稳定的. 这表明, 在约束的正则性与二阶充分性条件 (4.94)(可推出二阶增长条件) 假设成立, 只可能期望建立度 $\alpha = \dfrac{1}{2}$ 的最优解的 Hölder 稳定性. 现在推导 (非扰动) 问题 (P) 的局部唯一的最优解 x_0 邻域内的最优解的 $\dfrac{1}{2}$ 度 Hölder 稳定性的充分条件. 考虑最优化问题

$$\min_{x\in\Omega} g(x), \tag{4.95}$$

其中 $g: X \to \mathbb{R}$, $\Omega \subset X$. 我们视上述问题为目标函数与可行集均发生扰动的问题 (4.77) 的一个扰动. 令 $S_0 \subset \Phi$ 是问题 (4.77) 的最优解集. 设 S_0 是非空的, 二阶增长条件 (4.79) 成立. 在下述命题中, 我们导出 (4.95) 的最优解与 S_0 间的距离的界值.

命题 4.37 设 S_0 是问题 (4.77) 的最优解集, 令 N 是 S_0 的邻域, 满足

(i) 二阶增长条件 (4.79) 在 N 上成立.

(ii) 函数 f 与 g 在 N 上分别是模为 η_1 与 η_2 的 Lipschitz 连续的.

(iii) 差函数 $g-f$ 在 N 上是模为 κ 的 Lipschitz 连续的.

则 (4.95) 的任何 ε 最优解 $\bar{x} \in N$ 满足

$$\operatorname{dist}(\bar{x}, S_0) \leqslant c^{-1}\kappa + 2\delta_1 + c^{-1/2}(\eta_1\delta_1 + \eta_2\delta_2)^{1/2} + c^{-1/2}\varepsilon^{1/2}, \tag{4.96}$$

其中 $\delta_1 := \sup\limits_{x\in\Omega\cap N} \operatorname{dist}(x, \Phi\cap N)$, $\delta_2 := \sup\limits_{x_0\in S_0} \operatorname{dist}(x_0, \Omega\cap N)$.

证明 令 $\gamma_1 > 0$, $\gamma_2 > 0$ 与 $z \in \Phi\cap N$, $x_0 \in S_0$ 满足

$$\|\bar{x} - z\| \leqslant \operatorname{dist}(\bar{x}, \Phi\cap N) + \gamma_1, \tag{4.97}$$

$$\|z - x_0\| \leqslant \operatorname{dist}(z, S_0) + \gamma_2. \tag{4.98}$$

① 命题 4.36 中没有用到 $t_n \downarrow 0$, 而 $O(\|u-u_0\|^2)$ 应该为 $O(\|u_n-u_0\|^2)$.

由 δ_1 的定义, 可得

$$\|\bar{x}-z\| \leqslant \delta_1+\gamma_1. \tag{4.99}$$

考虑差函数 $h:=g-f$, $y\in\Omega\cap N$. 注意到 $\bar{x}\in\Omega\cap N$, 因而 $\Omega\cap N$ 是非空的. 得 $g(\bar{x})\leqslant g(y)+\varepsilon$, 有

$$\begin{aligned} f(\bar{x})-f(x_0)&=h(x_0)-h(\bar{x})+g(\bar{x})-g(y)+g(y)-g(x_0)\\ &\leqslant \kappa\|\bar{x}-x_0\|+\varepsilon+\eta_2\|y-x_0\|. \end{aligned}$$

因为 y 是 $\Omega\cap N$ 中的任意元素, 得到

$$f(\bar{x})-f(x_0)\leqslant \kappa\|\bar{x}-x_0\|+\varepsilon+\eta_2\delta_2. \tag{4.100}$$

另一方面, 由 $|f(\bar{x})-f(z)|\leqslant \eta_1\|\bar{x}-z\|$, 得到

$$\begin{aligned} f(\bar{x})-f(x_0)&\geqslant f(z)-f(x_0)-|f(\bar{x})-f(z)|\\ &\geqslant f(z)-f(x_0)-\eta_1\|\bar{x}-z\|. \end{aligned}$$

因为

$$\|z-x_0\|\geqslant \big|\,\|z-\bar{x}\|-\|\bar{x}-x_0\|\,\big|,$$

由二阶增长条件与 (4.98) 与 (4.99), 对小于上式右端的 $\gamma_2>0$ 有

$$\begin{aligned} f(\bar{x})-f(x_0)&\geqslant c(\|z-x_0\|-\gamma_2)^2-\eta_1\|\bar{x}-z\|\\ &\geqslant c(\|\bar{x}-x_0\|-\|z-\bar{x}\|-\gamma_2)^2-\eta_1(\delta_1+\gamma_1), \end{aligned} \tag{4.101}$$

因此对 $\gamma_1>0$ 与 $\gamma_2>0$ 充分小,

$$f(\bar{x})-f(x_0)\geqslant c(\|\bar{x}-x_0\|-\delta_1-\gamma_1-\gamma_2)^2-\eta_1(\delta_1+\gamma_1). \tag{4.102}$$

由 (4.100) 与 (4.102) 可得

$$\kappa\|\bar{x}-x_0\|+\beta_1\geqslant c(\|x-\bar{x}\|-\beta_2)^2-\beta_3, \tag{4.103}$$

其中 $\beta_1:=\varepsilon+\eta_2\delta_2$, $\beta_2:=\delta_1+\gamma_1+\gamma_2$ 及 $\beta_3:=\eta_1(\delta_1+\gamma_1)$. 解此二次不等式, 可得

$$\|\bar{x}-x_0\|\leqslant \frac{2c\beta_2+\kappa+\sqrt{(2c\beta_2+\kappa)^2-4c(c\beta_2^2-\beta_3-\beta_1)}}{2c}.$$

因为 $\mathrm{dist}(\bar{x},S_0)\leqslant\|\bar{x}-x_0\|$, 对 $a\geqslant 0$, $b\geqslant 0$, 有 $\sqrt{a^2+b^2}\leqslant a+b$, 于是有

$$\mathrm{dist}(\bar{x},S_0)\leqslant \frac{2c\beta_2+\kappa+(2c\beta_2+\kappa)+\sqrt{4c(\beta_3+\beta_1)}}{2c}.$$

由于正常数 γ_1 与 γ_2 可取任意小, 上述不等式可推出估计式 (4.96). □

注 4.38 若差函数 $g-f$ 的 Lipschitz 常数 κ 不 "小", 可以用差函数绝对值的上确界来代替相应的估计. 令

$$\sigma := \sup_{x\in N} \mid g(x)-f(x) \mid .$$

则 $|h(x_0)|$ 与 $|h(\bar{x})|$ 小于或等于 σ, 从而不等式 (4.100) 可由下式代替

$$f(\bar{x})-f(x_0) \leqslant 2\sigma+\varepsilon+\eta_2\delta_2.$$

与 (4.102) 相结合, 这可推出

$$2\sigma+\varepsilon+\eta_1\delta_1+\eta_2\delta_2 \geqslant c(\|\bar{x}-x_0\|-\delta_1)^2,$$

从而

$$\operatorname{dist}(\bar{x}, S_0) \leqslant c^{-1/2}(2\sigma)^{1/2}+c^{-1/2}\varepsilon^{1/2}+c^{-1/2}(\eta_1\delta_1+\eta_2\delta_2)^{1/2}+\delta_1. \tag{4.104}$$

尤其, 若 $\Phi=\Omega$, 有 $\delta_1=\delta_2=0$, 得到

$$\operatorname{dist}(\bar{x}, S_0) \leqslant c^{-1/2}(2\sigma)^{1/2}+c^{-1/2}\varepsilon^{1/2}. \tag{4.105}$$

这给出距离 $\operatorname{dist}(\bar{x}, S_0)$ 阶为 $O(\sigma^{1/2})$ 的一估计. 可以把这一估计与用 Lipschitz 常数 κ 表述的估计 (4.80) 相比较.

注 4.39 在命题 4.37 中的假设 (ii) 与 (iii) 成立的情况下, 若一阶增长性条件成立, 可以重复上述证明中的推导, 将不等式 (4.101) 与 (4.103) 用

$$\begin{aligned} f(\bar{x})-f(x_0) &\geqslant c(\|z-x_0\|-\gamma_2)-\eta_1\|\bar{x}-z\| \\ &\geqslant c(\|\bar{x}-x_0\|-\delta_1-\gamma_1-\gamma_2)-\eta_1(\delta_1+\gamma_1) \end{aligned}$$

与

$$\kappa\|\bar{x}-x_0\|+\varepsilon+\eta_2\delta_2 \geqslant c(\|\bar{x}-x_0\|-\delta_1-\gamma_1-\gamma_2)-\eta_1(\delta_1+\gamma_1)$$

代替. 若 $\kappa<c$, 这可导致下述估计

$$\operatorname{dist}(\bar{x}, S_0) \leqslant \frac{(c+\eta_1)\delta_1+\eta_2\delta_2+\varepsilon}{c-\kappa}. \tag{4.106}$$

即在一阶增长条件成立的前提下, 假如 Lipschitz 常数 κ 充分小, 则距离 $\operatorname{dist}(\bar{x}, S_0)$ 以距离 δ_1 与 δ_2 的线性函数为上界.

注 4.40 两个集合 $S,\ T\subset X$ 间的 Hausdorff 距离定义为

$$\operatorname{Haus}(S,T) := \max\left\{\sup_{x\in S}\ \operatorname{dist}(x,T),\ \sup_{x\in T}\ \operatorname{dist}(x,S)\right\}. \tag{4.107}$$

则可得 $\delta_1 \leqslant \rho$, $\delta_2 \leqslant \rho$, 其中

$$\rho := \text{Haus}(\Omega \cap N,\ \Phi \cap N), \tag{4.108}$$

这里 δ_1 与 δ_2 是上述命题中定义的距离. 所以, (4.96) 与 (4.106) 估计中的常数 δ_1 与 δ_2 可替换为 Hausdorff 距离常数 ρ.

界估计 (4.80) 与 (4.96) 尤其适宜, 因为它们由数据, 即 (4.80) 中的差函数的 Lipschitz 常数及 (4.96) 中的可行集间的距离的扰动给出. 将最优化问题 (P_{u_0}) 与 (P_u) 分别等同于问题 (4.77) 与 (4.95), 则这些界值结果可以应用.

命题 4.41 设 $\varepsilon(u)$ 是 $O(\|u-u_0\|)$ 阶的, 设

(i) 非扰动问题 (P_{u_0}) 具有非空的紧致的最优解集 S_0.

(ii) (对非扰动问题的) 二阶增长条件成立.

(iii) $f(x,u)$ 是二次连续可微的, $G(x,u)$ 是连续可微的.

(iv) 在 S_0 中的每一点处 Robinson 约束规范成立.

则存在 S_0 的邻域 N 及常数 $\alpha \geqslant 0$, 对充分接近于 u_0 的所有的 u, 若 $\bar{x}(u)$ 是 (P_u) 的 $\varepsilon(u)$ 最优解且 $\bar{x}(u) \in N$, 则

$$\text{dist}(\bar{x}(u), S_0) \leqslant \alpha \|u-u_0\|^{1/2}. \tag{4.109}$$

证明 由定理 2.87, 对每一 $x_0 \in S_0$, 存在 x_0 的一个邻域 $N(x_0)$ 满足 $\Phi(u) \cap N(x_0)$ 与 $\Phi(u_0) \cap N(x_0)$ 间的 Hausdorff 距离是 $O(\|u-u_0\|)$ 阶的. 因为 S_0 是紧致的, 则存在 S_0 的邻域 N 满足 $\Phi(u) \cap N$ 与 $\Phi(u_0) \cap N$ 间的 Hausdorff 距离是 $O(\|u-u_0\|)$ 阶的. 我们还有差函数 $f(\cdot,u) - f(\cdot,u_0)$ 在 S_0 的一邻域上是 $O(\|u-u_0\|)$ 阶的, 对于接近于 u_0 的所有的 u, $f(\cdot,u)$ 在此邻域上的 Lipschitz 常数是有界的. 则由 (4.96) 可得到 (4.109). □

当然, 由 (4.109) 可得, 当 $u \to u_0$ 时, (P_u) 的任何 $O(\|u-u_0\|)$ 最优解 $\bar{x}(u)$ 趋于 S_0, 有

$$\text{dist}(\bar{x}(u), S_0) = O(\|u-u_0\|^{1/2}). \tag{4.110}$$

注 4.42 若代替上述命题中的条件 (ii)(二阶增长条件), 假设一阶增长条件成立, 则由估计式 (4.106) 可得, 对任何 (P_u) 中的 $O(\|u-u_0\|)$ 最优解 $\bar{x}(u)$, 当 $u \to u_0$ 时, 它趋于 S_0, 有

$$\text{dist}(\bar{x}(u), S_0) = O(\|u-u_0\|). \tag{4.111}$$

4.4.3 Lagrange 乘子的定量稳定性

设可行集由 (4.2) 形式的抽象约束定义, 其中 $f(x,u)$ 与 $G(x,u)$ 是二阶连续可微的, 考虑相应的 Lagrange 乘子的连续性质. 由 Ekeland 变分原理 (定理 3.22), 若

$\bar{x}$ 是 (P_u) 的 ε 最优解, Robinson 约束规范在 $\overline{B}(\bar{x}, \varepsilon^{1/2})$ 中的任何可行点处均是成立的, 则对充分小的 $\varepsilon \geqslant 0$ 与 $\delta := \varepsilon^{1/2}$, 存在 (P_u) 的 ε 最优解 $\hat{x}$ 及 $\lambda \in Y^*$ 满足 $\|\hat{x} - \bar{x}\| \leqslant \delta$ 且

$$\|D_x L(\hat{x}, \lambda, u)\| \leqslant \delta, \quad G(\hat{x}, u) \in K, \quad \lambda \in N_K(G(\hat{x}, u)) \tag{4.112}$$

(见定理 3.23). 用 $\Lambda_\delta(\hat{x}, u)$ 记满足 (4.112) 的 $\lambda \in Y^*$ 的集合. 当然, 当 $\delta = 0$ 时, 上述集合简化为 Lagrange 乘子的集合.

命题 4.43 设 Robinson 约束规范在 (x_0, u_0) 处成立. 则对任何 $\delta \geqslant 0$, 集合 $\Lambda_\delta(\hat{x}, u)$ 对 (x_0, u_0) 的某一邻域中所有的 $(\hat{x}, u)$ 是一致有界的.

证明 利用广义凸映射定理 (定理 2.70), 由 Robinson 约束规范 (2.163) 有, 存在 $\varepsilon > 0$ 满足

$$\varepsilon B_Y \subset G(x_0, u_0) + D_x G(x_0, u_0) B_X - K.$$

令 $\lambda \in \Lambda_\delta(\hat{x}, u)$, 令 $y \in B_Y$ 满足 $\langle \lambda, y \rangle \geqslant \frac{1}{2}\|\lambda\|$. 则可以写为

$$-\varepsilon y = G(x_0, u_0) + D_x G(x_0, u_0) h - k$$

对某一 $h \in B_X$ 与一 $k \in K$ 成立. 可得到

$$\varepsilon y = -D_x G(\hat{x}, u) h + k - G(\hat{x}, u) + r,$$

其中

$$\|r\| \leqslant \|G(\hat{x}, u) - G(x_0, u_0)\| + \|D_x G(\hat{x}, u) - D_x G(x_0, u_0)\|.$$

因为 $\lambda \in N_K(G(\hat{x}, u))$, 有 $\langle \lambda, k - G(\hat{x}, u) \rangle \leqslant 0$, 得

$$\frac{1}{2}\varepsilon\|\lambda\| \leqslant \langle \lambda, \varepsilon y \rangle \leqslant -\langle \lambda, D_x G(\hat{x}, u) h \rangle + \|\lambda\| \, \|r\|.$$

进一步, 由 $G(x, u)$ 与 $D_x G(x, u)$ 的连续性, 可以选取 (x_0, u_0) 的某一邻域满足 $\|r\| \leqslant \frac{1}{4}\varepsilon$, 因为

$$-\langle \lambda, D_x G(\hat{x}, u) h \rangle \leqslant \langle D_x f(\hat{x}, u), h \rangle + \delta,$$

得到 $\frac{1}{4}\varepsilon\|\lambda\| \leqslant \|D_x f(\hat{x}, u)\| + \delta$. 由连续性性质, $D_x f(\hat{x}, u)$ 在 (x_0, u_0) 附近是有界的, 这就证得结论. □

可以将最优性条件 (4.112) 写为下述形式

$$(D_x L(x, \lambda, u), G(x, u)) \in \delta B_{X^*} \times N_K^{-1}(\lambda), \tag{4.113}$$

其中 N_K^{-1} 是多值函数 $y \mapsto N_K(y)$ 的图逆 (即 $y \in N_K^{-1}(\lambda)$ 当且仅当 $\lambda \in N_K(y)$). 考虑 Banach 空间 $Z := X \times Y^*$, $W := X^* \times Y$, 定义映射 $F : Z \times U \to W$ 与多值函数 $\Gamma : Z \to 2^W$:

$$F(z,u) := (D_x L(x,\lambda,u), -G(x,u)), \quad \Gamma(z) := (\delta B_{X^*}, N_K^{-1}(\lambda)), \tag{4.114}$$

其中 $z := (x, \lambda)$. 则一阶最优性条件 (4.112) 可以表述为下述形式

$$0 \in F(z,u) + \Gamma(z). \tag{4.115}$$

称 (4.115) 为广义方程 (5.12 节中将进一步讨论广义方程). 注意, 若对偶 Banach 空间赋予弱 * 拓扑, 则上述定义的 Banach 空间 Z 与 W 是彼此为对偶的.

我们可引入一条件, 它可视为 Lagrange 乘子的稳定性条件. 考虑从 $X^* \times Y$ 到 2^{Y^*} 的如下定义的多值函数

$$\Omega(a,y) := \{\lambda \in Y^* : [D_x G(x_0,u_0)]^* \lambda + a = 0, \ \lambda \in N_K(y)\}. \tag{4.116}$$

显然, $\Omega(a_0, y_0) = \Lambda(x_0, u_0)$, 其中 $a_0 := D_x f(x_0,u_0)$, $y_0 := G(x_0,u_0)$. 讨论 Ω 在点 (a_0, y_0) 处是上 Lipschitz 连续的多值函数的情况. 即存在 $c > 0$ 满足

$$\Omega(a,y) \subset \Omega(a_0,y_0) + c(\|a - a_0\| + \|y - y_0\|) B_{Y^*}, \tag{4.117}$$

对接近于 $(a_0,y_0) = (D_x f(x_0,u_0), G(x_0,u_0))$ 所有的 (a,y) 均成立. 由 Ω 的上 Lipschitz 的连续性, 结合 Robinson 约束规范, 给出 Lagrange 乘子集合是上 Lipschitz 连续的充分条件.

引理 4.44 *设 Robinson 约束规范在 (x_0,u_0) 处成立, $\Lambda(x_0,u_0)$ 是非空的, $\Omega(a,y)$ 在 (a_0,y_0) 处是上 Lipschitz 连续的, 其中 $a_0 := D_x f(x_0,u_0)$, $y_0 := G(x_0,u_0)$. 则多值函数 $X \times U \times \mathbb{R}_+ \ni (x,u,\delta) \mapsto \Lambda_\delta(x,u)$ 在 $(x_0,u_0,0)$ 处是上 Lipschitz 连续的.*

证明 由条件 (4.117) 可推出, 对任何 $\lambda \in \Lambda_\delta(\hat{x},u)$,

$$\begin{aligned} \operatorname{dist}(\lambda, \Lambda(x_0,u_0)) \leqslant c\ (&\|D_x f(\hat{x},u) - D_x f(x_0,u_0)\| \\ &+ \|\lambda\|\ \|D_x G(\hat{x},u) - D_x G(x_0,u_0)\| \\ &+ \|G(\hat{x},u) - G(x_0,u_0)\| + \delta). \end{aligned}$$

因为 $\Lambda_\delta(x_0,u_0)$①在 (x_0,u_0) 与 $\delta = 0$ 的邻域内是一致有界的 ②, 由前面的不等式得 $\Lambda_\delta(\hat{x},u)$ 在 (x_0,u_0) 的附近是一致有界的, 且若 $\Lambda_\delta(\hat{x},u) \neq \varnothing$, 则

$$\sup_{\lambda \in \Lambda_\delta(\hat{x},u)} \operatorname{dist}(\lambda, \Lambda(x_0,u_0)) = O(\|\hat{x} - x_0\| + \|u - u_0\| + \delta). \tag{4.118}$$

① 应为 $\Lambda_\delta(x,u)$.

② 或 $\Lambda_\delta(x_0,u_0)$ 在 $\delta = 0$ 处是一致有界的.

这可证得结论. □

注 4.45 (i) 多值函数 Ω 在 (a_0, y_0) 处是上 Lipschitz 连续的充分条件是 $D_xG(x_0, u_0)$ 是映上的, (ii) 另一充分条件是集合 K 为凸的广义多面集 (见 2.5.7 节). 事实上, 此种情况, 法锥多值函数 $y \mapsto N_K(y)$ 是广义多面集, 尽管它可能是非凸的 (见例 2.209). 由于法锥多值函数的图的每一凸广义多面组分与仿射空间的交还是一凸的广义多面体集合, 因此 Ω 的图是广义的多面体. 则由定理 2.208 可得, 映射 Ω 是上 Lipschitz 连续的.

下面介绍 (4.117) 的另一 (相对强的) 充分条件, 由上述引理, 可推出 Lagrange 乘子的上 Lipschitz 连续性.

定义 4.46 *称 Lagrange 乘子 $\lambda_0 \in \Lambda_0(x_0, u_0)$ 满足严格约束规范, 若*

$$0 \in \operatorname{int}\{y_0 + D_xG(x_0, u_0)X - K_0\}, \tag{4.119}$$

其中 $y_0 := G(x_0, u_0)$ 且

$$K_0 := \{y \in K : \langle \lambda_0, y - y_0 \rangle = 0\}. \tag{4.120}$$

条件 (4.119) 即相对于简约集合 K_0 的 Robinson 约束规范. 因为 $K_0 \subset K$, 严格约束规范可推出 Robinson 约束规范. 注意, 若 K 是凸锥, 则 (由一阶最优性条件)$\langle \lambda_0, y_0 \rangle = 0$, 因此有 $K_0 = K \cap \ker\lambda_0$.

严格约束规范可推出的结论将在下面的命题中给出. 为简化记号, 记 $y_0 := G(x_0, u_0)$, $\Lambda_0 := \Lambda(x_0, u_0)$

命题 4.47 *设 λ_0 满足一阶最优性条件 (3.16) 且严格约束规范 (4.119) 成立. 则*

(i) Lagrange 乘子向量 λ_0 是唯一的, 即 $\Lambda_0 = \{\lambda_0\}$.

(ii) 由 (4.116) 定义的多值函数 Ω 在 $(D_xf(x_0, u_0), y_0)$ 处是上 Lipschitz 连续的.

相反地, 令

$$\mathcal{Q}(x_0) := D_xG(x_0, u_0)X + \mathcal{R}_{K_0}(y_0). \tag{4.121}$$

若 $\Lambda_0 = \{\lambda_0\}$, 且锥 $\mathcal{Q}(x_0)$ 与 $\mathcal{R}_{N_K(y_0)}(\lambda_0)$ 是闭的, 则 (4.119) 成立.

证明 令 $\lambda \in \Lambda_0$. 置 $A := D_xG(x_0, u_0)$, $\mu := \lambda - \lambda_0$. 因为 λ 与 λ_0 是 Lagrange 乘子, 由一阶最优性条件有 $A^*\lambda_0 = A^*\lambda$, 从而对所有的 $h \in X$, $\langle \mu, Ah \rangle = \langle A^*\mu, h \rangle = 0$. 由 (4.119), 对每一 $y \in Y$, 存在 $\varepsilon > 0$, $h \in X$ 及 $\kappa_0 \in K_0$ 满足 $\varepsilon y = y_0 + Ah - \kappa_0$. 因为 $\langle \mu, Ah \rangle = 0$, $\lambda \in N_K(y_0)$, 因而得到

$$\langle \mu, y_0 - \kappa_0 \rangle = \langle \lambda, y_0 - \kappa_0 \rangle \geqslant 0,$$

得到 $\langle \mu, y \rangle \geqslant 0$. 因为 y 是 Y 中的任意元素, 可推出 $\mu = 0$. 证得 (i).

因为 λ_0 是唯一的, 有 $\Omega(a_0,y_0)=\{\lambda_0\}$, 其中 $a_0:=D_xf(x_0,u_0)$, $y_0:=G(x_0,u_0)$. 考虑 $\lambda\in\Omega(a,y)$. 置 $\eta:=\lambda_0-\lambda, z:=y-y_0, b:=a-a_0$, 令 $\bar{y}\in B_Y$ 满足 $\frac{1}{2}\|\eta\|\leqslant-\langle\eta,\bar{y}\rangle$. 根据广义开映射定理(定理 2.70), 由 (4.119), 存在 $\varepsilon>0$ 满足

$$\varepsilon B_Y\subset AB_X-(K_0-y_0). \tag{4.122}$$

因此, 存在 $h\in B_X$, $\kappa_0\in K_0$ 满足 $\varepsilon\bar{y}=Ah-\kappa_0+y_0$. 因为 $\lambda\in N_K(y_0+z)$, 有 $\langle\lambda,\kappa_0-y_0-z\rangle\leqslant 0$, 所以有

$$-\langle\eta,\kappa_0-y_0\rangle=\langle\lambda,\kappa_0-y_0\rangle\leqslant\langle\lambda,z\rangle\leqslant(\|\eta\|+\|\lambda_0\|)\|z\|.$$

将此与 $A^*\eta=-b$ 结合, 得到

$$\frac{1}{2}\varepsilon\|\eta\|\leqslant-\langle\eta,\varepsilon\bar{y}\rangle=-\langle\eta,Ah\rangle+\langle\eta,\kappa_0-y_0\rangle\leqslant\|b\|+(\|\eta\|+\|\lambda_0\|)\|z\|.$$

当 $\|y-y_0\|\leqslant\varepsilon/4$ 时, 有

$$\|\lambda-\lambda_0\|\leqslant 4\varepsilon^{-1}(\|a-a_0\|+\|\lambda_0\|\ \|y-y_0\|), \tag{4.123}$$

于是证得 (ii).

现在设 $\Lambda_0=\{\lambda_0\}$ 且锥 $\mathcal{Q}(x_0)$ 与 $\mathcal{R}_{N_K(y_0)}(\lambda_0)$ 是闭的, 但条件 (4.119) 不成立. 则由命题 2.95, 凸闭锥 $\mathcal{Q}(x_0)$ 不等于 Y. 因此极锥 $\mathcal{Q}(x_0)^-$ 含非零的元素 ξ. 由 $\mathcal{Q}(x_0)$ 的表示可得①

$$\mathcal{Q}(x_0)^-=[DG(x_0)X]^\perp\cap N_{K_0}(y_0).$$

所以有 $[DG(x_0)]^*\xi=0$ 且 $\xi\in N_{K_0}(y_0)$.

因为 $K_0-y_0=(K-y_0)\cap\ker\lambda_0$, 有 $\mathcal{R}_{K_0}(y_0)=\mathcal{R}_K(y_0)\cap\ker\lambda_0$, 因此, 由 (2.32) 得

$$N_{K_0}(y_0)=\mathrm{cl}\{N_K(y_0)+[\![\lambda_0]\!]\}.$$

另一方面, 由 Λ_0 的定义与 $\mathcal{R}_{N_K(y_0)}(\lambda_0)$ 的闭性, 可得 (见例 2.62)

$$\begin{aligned}\mathcal{R}_{\Lambda_0}(\lambda_0)&=\mathcal{R}_{N_K(y_0)}(\lambda_0)\cap[DG(x_0)X]^\perp\\&=T_{N_K(y_0)}(\lambda_0)\cap[DG(x_0)X]^\perp\\&=\mathrm{cl}(N_K(y_0)+[\![\lambda_0]\!])\cap[DG(x_0)X]^\perp.\end{aligned}$$

由 $\xi\in\mathcal{R}_{\Lambda_0}(\lambda_0)$, 对 $t>0$ 充分小, 有 $\lambda_0+t\xi\in\Lambda_0$. 这与 λ_0 的唯一性矛盾. □

注 4.48 若 (4.122) 中的常数 $\varepsilon>0$ 可选取与 y_0 无关, 则 Lagrange 乘子之变化的上界 (4.123) 在点 (a_0,y_0) 的一邻域内是一致的. 例如 $D_xG(x_0,u_0)$ 是映上的, 这一条件是成立的.

① 下式中 $DG(x_0)$ 应为 $D_xG(x_0,u_0)$.

注 4.49 上述命题表明, 关于简约集 K_0 的 Robinson 约束规范 (即严格约束规范) 是 Lagrange 乘子唯一性的充分条件. 在如 (2.174) 定义的有限多个约束的情况, K 是有限维的多面凸锥, 上述命题中后一结论所考虑的两个锥是多面的、闭的. 所以, 此种情况, 相对于 K_0 的 Robinson 约束规范是 λ_0 的唯一性的充分与必要条件. 此种情况, 由 Robinson 约束规范的刻画, 得此条件的等价形式是

$$\begin{aligned} &Dg_i(x_0),\, i \in \{1, \cdots, q\} \cup I_+(\lambda_0)\text{是线性无关的},\\ &\exists h \in X : Dg_i(x_0)h = 0, \quad i \in \{1, \cdots, q\} \cup I_+(\lambda_0),\\ &\qquad\quad Dg_i(x_0)h < 0, \quad i \in I(x_0) \setminus I_+(\lambda_0), \end{aligned} \tag{4.124}$$

其中 $I(x_0)$ 记不等式约束在 x_0 处起作用的指标集合, $I_+(\lambda_0)$ 记对应 λ_0 的非零 (正) 分量的不等式约束的指标集. 条件 (4.124) 等价于有限多个约束的情况下的严格约束规范.

现在考虑下述条件. 令 $\lambda_0 \in \Lambda(x_0, u_0)$, 设

$$D_xG(x_0, u_0)X - T_K(y_0) \cap \ker \lambda_0 = Y, \tag{4.125}$$

其中 $y_0 := G(x_0, u_0)$. 条件 (4.125) 是严格约束规范 (4.119) 应用于映射 $D_xG(x_0, u_0)$ 与锥 $T_K(y_0)$ 的特殊情况. 还注意, 因为 $\mathcal{R}_{K_0}(y_0) \subset T_K(y_0) \cap \ker \lambda_0$, 条件 (4.125) 可由条件 (4.119) 推得. 下述命题表明, 上述条件 (4.125) 确保对任何方向 d, 线性化问题(PL_d) 具有最优解.

命题 4.50 令 $\lambda_0 \in \Lambda(x_0, u_0)$, 设条件 (4.125) 成立. 则

(i) Lagrange 乘子 λ_0 是唯一的, 即 $\Lambda(x_0, u_0) = \{\lambda_0\}$,

(ii) 对任何 d, 问题 (PL_d) 具有最优解 $\bar{h}(d)$ 满足 $\bar{h}(d) = O(\|d\|)$.

证明 结论 (i) 是命题 4.47 中相应结论的特殊情况. 为证明 (ii), 我们如下进行. 由 (4.125), 存在 $\bar{h} \in X$ 满足 $DG(x_0, u_0)(\bar{h}, d) \in Q_0$, 其中 $Q_0 := T_K(y_0) \cap \ker \lambda_0$, 因此 $\bar{h}$ 是 (PL_d) 的可行点. 由锥 Q_0 的定义, 有 $\langle \lambda_0, DG(x_0, u_0)(\bar{h}, d)\rangle = 0$, 因为 $\lambda_0 \in \Lambda(x_0, u_0)$, 得 $\bar{h}$ 与 λ_0 满足问题 (PL_d) 的一阶最优性条件, 即 $\bar{h}$ 是 (PL_d) 的最优解, λ_0 是 (DL_d) 的最优解. 剩下需证 $\bar{h}$ 可选取为 $O(\|d\|)$ 阶的. 为证明此结果, 考虑多值函数 $\mathcal{M}(h) := D_xG(x_0, u_0)h - Q_0$. 由 (4.125) 得 $0 \in \mathrm{int}(\mathrm{range}\mathcal{M})$, 从而由广义开映射定理有 $0 \in \mathrm{int}\{\mathcal{M}(B_X)\}$. 这就推得结论. □

后面我们将看到 (见例 4.54), 可能发生 Lagrange 乘子是唯一的但条件 (4.119) 与 (4.125) 不成立的情况.

4.4.4 最优解与 Lagrange 乘子的 Lipschitz 稳定性

这一节研究满足一阶最优性条件的最优解与 Lagrange 乘子的稳定性. 称点 $x \in X$ 是最优化问题的稳定点, 若相应的 Lagrange 乘子集合非空. 下述定理表明, 若在

(4.116) 中定义的 Ω 多值函数在点 $(D_x f(x_0, u_0), y_0)$ 处是上 Lipschitz 连续的, 约束是正则的, 且强的二阶充分性条件成立, 则最优解与相联系 Lagrange 乘子 (关于扰动) 是上 Lipschitz 稳定的.

定理 4.51 令 $\hat{z}(u) = (\hat{x}(u), \hat{\lambda}(u))$ 是满足近似的最优性条件 (4.112) 的点对, 其中 $\delta = O(\|u - u_0\|)$, 设 $u \to u_0$ 时 $\hat{x}(u) \to x_0$. 设

(i) Robinson 约束规范 (2.163) 在 (x_0, u_0) 处成立.

(ii) (4.116) 中定义的多值函数 Ω 在点 $(D_x f(x_0, u_0), G(x_0, u_0))$ 处是上 Lipschitz 连续的.

(iii) 下述二阶条件成立: 存在 $\beta > 0$ 与 $\eta > 0$ 满足

$$D^2_{xx} L(x_0, \lambda, u_0)(h, h) \geqslant \beta \|h\|^2, \quad \forall \lambda \in \Lambda(x_0, u_0), \quad \forall h \in \widehat{C}_\eta(x_0), \tag{4.126}$$

其中近似临界锥 $\widehat{C}_\eta(x_0)$ 由 (3.147) 定义. 则

$$\|\hat{x}(u) - x_0\| + \operatorname{dist}(\hat{\lambda}(u), \Lambda(x_0, u_0)) = O(\|u - u_0\|). \tag{4.127}$$

证明 注意, 由定理 3.63, 二阶充分条件 (4.126) 可推出 x_0 是非扰动问题 (P) 的局部最优解. 在 Robinson 约束规范之下, $\Lambda(x_0, u_0)$ 是非空的, 由命题 4.43, $\Lambda_\delta(\hat{x}, u)$ 对 (x_0, u_0) 的邻域中所有的 $(\hat{x}, u)$ 是一致有界的, 因此, 假设 (ii) 可推出 Lagrange 乘子稳定性的估计 (4.118), 则 (4.127) 成立当且仅当 $\hat{x}(u)$ 在 x_0 处是 Lipschitz 稳定的.

用反证法. 若 (4.127) 不真, $\hat{x}(u)$ 在 x_0 处不是 Lipschitz 稳定的. 则存在 $u_n \to u_0$, $x_n := \hat{x}(u_n)$, $\lambda_n := \hat{\lambda}(u_n)$ 满足 (4.112), 其中 $\delta := \delta_n = O(\|u_n - u_0\|)$, 这里 $\delta_n := \|D_x L(x_n, \lambda_n, u_n)\|$, $x_n \to x_0$, 但有 $\lim_n \|x_n - x_0\| / \|u_n - u_0\| = \infty$. 置

$$\begin{aligned} \kappa_n &:= \|x_n - x_0\|, \quad & h_n &:= \kappa_n^{-1}(x_n - x_0), \\ \tau_n &:= \|u_n - u_0\|, \quad & z_n &:= (x_n, u_n). \end{aligned}$$

因为 $\Lambda_\delta(\hat{x}, u)$ 是一致有界的, 可得 $\{\lambda_n\}$ 是有界的.

对任意的 $\eta' > 0$, 由 f 与 G 的一阶展式可得, 对充分大的 n 有 $h_n \in C_{\eta'}(x_0)$. 将定理 2.87 应用于线性映射 $D_x G(x_0, u_0)$ 与集合 $T_k(G(x_0, u_0))$ 可得, 存在 h'_n 满足 $h'_n - h_n \to 0$ 且

$$D_x G(x_0, u_0) h'_n \in T_k(G(x_0, u_0)). \tag{4.128}$$

则对充分大的 n, $h'_n \in \widehat{C}_{\eta'}(x_0)$, 其中, η 是使 (4.126) 成立的常数. 令 $\lambda'_n \in \Lambda(x_0, u_0)$ 满足

$$\|\lambda_n - \lambda'_n\| = O(\|x_n, x_0\| + \tau_n). \tag{4.129}$$

多值函数 Ω 的上 Lipschitz 连续函数性及 $\delta_n = O(\tau_n)$ 可确保 λ'_n 的存在性 (见 (4.118)). 置 $z'_n := (x_0, \lambda'_n)$. 考虑对应于最优性条件 (4.112) 的广义方程 (4.115). 由于多值函数 $N_K(\cdot)$ 是单调的, 有

$$\langle \lambda_n - \lambda'_n, G(x_n - u_n) - G(x_0, u_0) \rangle \geqslant 0. \tag{4.130}$$

对由 (4.114) 定义的 F, 有

$$\begin{aligned}\Delta_1 &:= \langle z_n - z'_n, F(z_n, u_n) - F(z'_n, u_0) \rangle \\ &\leqslant \langle x_n - x_0, D_x L(x_n, \lambda_n, u_n) - D_x L(x_0, \lambda'_n, u_0) \rangle \\ &= \langle x_n - x_0, D_x L(x_n, \lambda_n, u_n) \rangle \leqslant \|x_n - x_0\| \delta_n = O(\kappa_n \tau_n).\end{aligned}$$

另一方面

$$D_z F(z, u_0) = \begin{bmatrix} D^2_{xx} L(x, \lambda, u_0) & D_x G(x, u_0) \\ -D_x G(x, u_0)^* & 0 \end{bmatrix}.$$

对任何 $w = (h, \lambda)$ 有

$$\langle w, D_z F(z, u_0) w \rangle = \langle h, D^2_{xx} L(x, \lambda, u_0) h \rangle.$$

因此, 由于 $\operatorname{dist}(h_n, C_\eta(x_0)) \to 0$, 对充分大的 n, 由 (4.126) 可得

$$\begin{aligned}\Delta_2 &:= \langle z_n - z'_n, F(z_n, u_0) - F(z'_n, u_0) \rangle \\ &= \left\langle z_n - z'_n, \int_0^1 D_z F(z'_n + t(z_n - z'_n), u_0)(z_n - z'_n) \mathrm{d}t \right\rangle \\ &= \kappa_n^2 \int_0^1 D^2_{xx} L(x_0 + t(x_n - x_0), \lambda'_n + t(\lambda_n - \lambda'_n), u_0)(h_n, h_n) \mathrm{d}t \\ &\geqslant \frac{1}{2} \beta \kappa_n^2.\end{aligned}$$

然而

$$\Delta_1 - \Delta_2 = \langle z_n - z'_n, F(z_n, u_n) - F(z_n, u_0) \rangle = O(\kappa_n \tau_n).$$

因为 $\Delta_1 = O(\kappa_n \tau_n), \Delta_2 \geqslant \dfrac{1}{2} \beta \kappa_n^2$ 且 $\kappa_n / \tau_n \to +\infty$, 这导致矛盾. □

与4.5节要讨论的最优解的方向 Lipschitz 稳定性的相应结果相比较, 二阶充分条件 (4.126) 是相当强的. 注意到, 若 Lagrange 乘子是唯一的, 则二阶充分条件 (4.126) 与条件 (4.138) 重合, 后者用于确保最优解的 Lipschitz 方向稳定性. 这里还需要多值函数 $\Omega(\cdot, \cdot)$ 的上 Lipschitz 稳定性. 另一方面, Lipschitz 估计 (4.127) 关于 u 是一致的, 它是对稳定解或满足最优性条件 (4.112) 的 “接近稳定” 解给出的.

在 (4.127) 中取 $u = u_0$ 得到, 在定理 4.51 的假设下, x_0 是非扰动问题 (P) 的局部唯一的稳定点, 若集 K 是广义多面集, 可以在较弱的条件下导出这一结果. 注意到, 二阶条件 (4.126) 是 x_0 处的二阶增长条件的充分性条件, 但非必要的 (见 3.3.1 节).

命题 4.52 *考虑非扰动问题 (P), 令 x_0 是 (P) 的稳定点, 设*

(i) K 是凸的广义多面锥.

(ii) Robinson 约束规范 (2.163) 在 x_0 处成立.

(iii) 二阶增长条件在 x_0 处成立.

则 x_0 是问题 (P) 的局部唯一的稳定点.

证明 注意, 由于 Robinson 约束规范成立, 相应的 Lagrange 乘子集合 $\Lambda(x_0)$ 非空有界. 用反证法. 设结论不真. 则存在 (P) 的稳定点序列 $x_n(x_n \neq x_0)$, 收敛于 x_0. 令 λ_n 是相应的 Lagrange 乘子. 由 Hoffman 引理, 有 $\mathrm{dist}(\lambda_n, \Lambda(x_0)) \to 0$. 进一步, 由于 K 是广义的多面体集合, 可视 (P) 的可行集由一个等式和有限多个不等式约束来定义, 因此 $\Lambda(x_0)$ 是紧致的. 如有必要可取一子列, 不妨设 λ_n 收敛于一点 $\lambda_0 \in \Lambda(x_0)$.

$f(x_n) = L(x_n, \lambda_n), f(x_0) = L(x_0, \lambda_0)$, 由二阶增长条件有, 存在 $c > 0$, 对充分大的 n,

$$L(x_n, \lambda_n) - L(x_0, \lambda_0) \geqslant c\|x_n - x_0\|^2.$$

还有

$$L(x_n, \lambda_n) = L(x_n, \lambda_0) + \langle \lambda_n - \lambda_0, G(x_n) \rangle$$

且 $\langle \lambda_n, G(x_n) \rangle = 0$. 进一步, 由于 K 是广义的多面集, 如有必要可选取子列, 不妨设在 x_n 处起作用的约束的指标集是常值的. 对应于 x_n 处非起作用约束的 λ_n 的分量是零. 取极限可以得到对应于这些约束的 λ_0 的分量也是零. 则对充分大的 n 有 $\langle \lambda_0, G(x_n) \rangle = 0$. 所以可得

$$L(x_n, \lambda_0) - L(x_0, \lambda_0) \geqslant c\|x_n - x_0\|^2$$

对充分大的 n 成立. 由

$$L(x_n, \lambda_0) - L(x_0, \lambda_0) = \frac{1}{2}\kappa_n^2 D_{xx}^2 L(x_0, \lambda_0)(h_n, h_n) + o(\kappa_n^2),$$

其中 $h_n := \kappa_n^{-1}(x_n - x_0), \kappa_n := \|x_n - x_0\|$, 则存在 $\beta > 0$, 对充分大的 n,

$$D_{xx}^2 L(x_0, \lambda_0)(h_n, h_n) \geqslant \beta. \tag{4.131}$$

考虑映射 $F(z) := (D_x L(x, \lambda), -G(x))$, 其中 $z := (x, \lambda)$. 有稳定点应满足广义方程 $0 \in F(z) + \Gamma(z)$, 其中 $\Gamma(z) := (0, N_K^{-1}(\lambda))$. 因为多值函数 $\Gamma(z)$ 是单调的, 有

$$\langle z_n - z_0, \quad F(z_n) - F(z_0) \rangle \leqslant 0, \tag{4.132}$$

其中 $z_n := (x_n, \lambda_n), z_0 := (x_0, \lambda_0)$. 另一方面

$$
\begin{aligned}
&\langle z_n - z_0,\ \ F(z_n) - F(z_0)\rangle \\
&= \langle x_n - x_0,\ \ D_x L(x_n, \lambda_n) - D_x L(x_0, \lambda_0)\rangle - \langle \lambda_n - \lambda_0,\ \ G(x_n) - G(x_0)\rangle \\
&= \langle x_n - x_0,\ D_x L(x_n, \lambda_0) - D_x L(x_0, \lambda_0)\rangle \\
&\quad -\langle \lambda_n - \lambda_0,\ \ G(x_n) - G(x_0) - D_x G(x_n)(x_n - x_0)\rangle \\
&= \kappa_n^2\ [D_{xx}^2 L(x_0, \lambda_0)(h_n, h_n) - \frac{1}{2}\langle \lambda_n - \lambda_0,\ \ D_{xx}^2 G(x_0)(h_n, h_n)\rangle] + o(\kappa_n^2) \\
&= \kappa_n^2 D_{xx}^2 L(x_0, \lambda_0)(h_n, h_n) + o(\kappa_n^2).
\end{aligned}
$$

与 (4.131) 相结合, 这可推出

$$
\langle z_n - z_0,\ \ F(z_n) - F(z_0)\rangle > 0
$$

对充分大的 n 成立, 这与 (4.132) 相矛盾. □

4.5 最优解的方向稳定性

这一节讨论最优 (接近最优) 解当参数向量 u 的扰动切于某一给定方向 $d \in U$ 时的量化稳定性. 这一节采用的技术基于沿方向 d 的最优值函数的上方、下方估计, 这在后继的方向分析中也是相当典型的方法. 这一节, 设 X, Y 与 U 是 Banach 空间, 可行集由 (4.2) 形式的抽象约束定义, 目标函数 $f(x,u)$ 与约束映射 $G(x,u)$ 是二次连续可微的. 将非扰动问题 (P_{u_0}) 视为 3.2.2 节与 3.3 节所讨论的问题 (P). 用在定义 4.8 给出的、在定理 4.9 中讨论的方向正则性条件. Robinson 约束规范可推出沿每一方向 $d \in U$ 的方向正则性.

最优值函数的下方估计依赖于在 3.3 节引入的应用于非扰动问题 (P) 的各种二阶充分性条件. 回顾 3.3 节中讨论的两种类型的二阶充分性条件. 第一种类型在定义 3.60 中给出, 不包括 "sigma 项", 这一项表示集合 K 的一种可能的曲率. 这种类型的条件在 3.2.3 节讨论的广义的多面性质之下成为 "几乎必要的". 第二种类型由定理 3.83 给出, 要求 X 是有限维的, 基于上二阶近似集的一个合适的选取. 尤其, 在二阶正则性 (见定义 3.85) 的假设之下, 可以取相应的 (内) 二阶切集作为上二阶近似集合, 此种情形, 二阶必要与充分最优性条件之间没有间隙.

4.5.1 Hölder 方向稳定性

我们首先讨论扰动切于给定的方向 $d \in U$, 沿下述形式

$$
u(t) := u_0 + td + o(t), \quad t \geqslant 0 \tag{4.133}
$$

的路径最优解的 $\frac{1}{2}$ 度 Hölder 稳定性.

定理 4.53 令 $x_0 \in \Phi(u_0)$ 满足在 x_0 点沿方向 $d \in U$ 的方向正则性条件成立. 令 $\bar{x}(t)$ 是当 $t \downarrow 0$ 收敛于 x_0 的 $(\mathrm{P}_{u(t)})$ 的 $O(t)$ 最优解. 进一步设或者 (i) 二阶充分条件 (3.136) 成立, 或者 (ii) 空间 X 是有限维的且定理 3.83 中的充分条件 (3.155) 对相应的二阶近似集族中某一选取 $\mathcal{A}(h)$ 成立. 则

$$\|\bar{x}(t) - x_0\| = O(t^{1/2}). \tag{4.134}$$

证明 由二阶充分条件可推出 x_0 是非扰动问题 (P) 的局部最优解. 因为 $\bar{x}(t)$ 收敛到 x_0, 如有必要, 可压缩可行域, 不妨设 x_0 是 (P) 的最优解. 用反证法. 若 (4.134) 不真, 则存在序列 $t_n \downarrow 0$ 满足, 若置

$$x_n := \bar{x}(t_n), \quad u_n := u(t_n), \quad \tau_n := \|x_n - x_0\|, \quad h_n := \tau_n^{-1}(x_n - x_0),$$

则 x_n 是 (P_{u_n}) 的 $O(t_n)$ 最优解, 但 $\tau_n^2/t_n \to \infty$. 由命题 4.21 与 4.22, 因为 $t_n = o(\tau_n^2)$, 有

$$f(x_n, u_n) - f(x_0, u_0) \leqslant v(u_n) - v(u_0) + O(t_n) \leqslant O(t_n) = o(\tau_n^2). \tag{4.135}$$

因为 $t_n = o(\tau_n^2)$, 由 f 与 G 的一阶展开式, 对任何 $\eta > 0$, 对充分大的 n, η_n 属于由 (3.36) 定义的近似临界锥$C_\eta(x_0)$.

我们将导出基于 (i) 或 (ii) 给出的二阶充分性条件成立时的下方近似. 这些推导分别类似于定理 3.63 与定理 3.83 的证明. 先设 (i) 成立. 令 $\gamma_n = (\alpha_n, \lambda_n) \in \Lambda_N^g(x_0)$ 是正规化的广义 Lagrange 乘子, 满足对充分大的 n, 有

$$D_{xx}^2 L^g(x_0, \gamma_n, u_0)(h_n, h_n) \geqslant \beta > 0. \tag{4.136}$$

考虑两种情况. 设对充分大的 n, 广义 Lagrange 乘子 (α_n, λ_n) 是非奇异的. 则 (如 (3.143)), 由于 $t_n = o(\tau_n^2), D_x L^g(x_0, \gamma_n, u_0) = 0$, 集合 $\Lambda_N^g(x_0)$ 是有界的, 有

$$\begin{aligned}\Delta :&= f(x_n, u_u) - f(x_0, u_0)\\ &\geqslant (\alpha_n)^{-1}(L^g(x_n, \gamma_n, u_n) - L^g(x_0, \gamma_n, u_0))\\ &= (\alpha_n)^{-1}\left[\frac{1}{2}\tau_n^2 D_{xx}^2 L^g(x_0, \gamma_n, u_0)(h_n, h_n) + o(\tau_n^2)\right] \geqslant \kappa\tau_n^2,\end{aligned}$$

其中 $\kappa > 0, n$ 充分大, 这与 (4.135) 矛盾.

再假设 $\alpha_n = 0$ 对充分大的 n 成立. 因为 $\lambda_n \in N_K(G(x_0, u_0)), D_x L^g(x_0, \gamma_n, u_0) = 0$, 对充分大的 n 有

$$\begin{aligned}0 &\geqslant \langle \lambda_n, \ G(x_n, u_n) - G(x_0, u_0)\rangle = L^g(x_n, \gamma_n, u_n) - L^g(x_0, \gamma_n, u_0)\\ &= \frac{1}{2}\tau_n^2 D_{xx}^2 L^g(x_0, \gamma_n, u_0)(h_n, h_n) + o(\tau_n^2) > 0,\end{aligned}$$

这导致矛盾.

设 (ii) 成立, 因为空间 X 是有限维的, 如有必要, 可取一子列, 存在 $h \in C(x_0)$ 满足 $h \neq 0, x_n = x_0 + \tau_n h + o(\tau_n), \tau_n \downarrow 0$. 类似于定理 3.83 证明中相应的推导, 只需将定理 3.83 证明中的 t_n 用 τ_n 代替, 即可完成本定理的证明. □

4.5.2 Lipschitz 方向稳定性

为建立最优解的 Lipschitz 稳定性, 需要加强由 (4.135) 给出的最优值函数的上估计. 如在 4.4.2 节中指出的那样, 定理 4.53 的二阶条件对最优解的 Lipschitz 稳定性不是充分的. 进一步, 当处理非多面凸集 K 时, 需要额外的假设. 下面的例子说明了这一点.

例 4.54 考虑 2×2 阶对称矩阵构成的线性空间 $Y := \mathcal{S}^2$, 锥 $K := \mathcal{S}^2_+$ 由 2×2 阶半正定矩阵构成. 考虑线性映射 $G : \mathbb{R}^2 \times \mathcal{S}^2 \to \mathcal{S}^2$, 由 $G(x_1, x_2, A) := \text{diag}(x_1, x_2) + A$ 定义, 其中 $\text{diag}(x_1, x_2)$ 记对角元素为 x_1 与 x_2 的对角矩阵, 参数化问题是

$$\min_{x\in\mathbb{R}^2} \; x_1 + x_1^2 + x_2^2 \quad \text{s. t.} \quad G(x_1, x_2, A) \in K. \tag{4.137}$$

对 $A = 0$, 相应的 (非扰动) 问题的可行集是 $\mathbb{R}^2_+$, 且其最优解是 $x_0 = (0, 0)$. 这时 Slater 条件成立, 非扰动问题有唯一的 Lagrange 乘子 $\lambda_0 = \text{diag}(-1, 0)$. 进一步, 问题 (4.137) 是凸的, 二阶最优性条件的一个较强的形式是成立的. 然而, 不难验证, 对任何非对角阵 $A \in \mathcal{S}^2$, 问题 $(P_{tA}), t \geqslant 0$ 具有唯一的最优解 $\bar{x}(t) = (\bar{x}_1(t), \bar{x}_2(t))$, 其中当 $t \downarrow 0$ 时, $\bar{x}_2(t)$ 是 $t^{2/3}$ 阶的. 因此, 最优解 $\bar{x}(t)$ 沿任何非对角阵方向 A 不是 Lipschitz 稳定的.

有意思的是, 在此例中锥 $K_0 := K \cap \ker \lambda_0$ 由形式为 $\text{diag}(0, a), a \geqslant 0$ 的对角矩阵组成, 尽管 λ_0 是唯一的, 因此相对于锥 K_0 的 Robinson 约束规范不成立 (与命题 4.47 的最后一个结论相比较).

回顾问题 (P_u) 的由 (4.45) 给出的线性化问题 (PL_d). 设 Lagrange 乘子集合 $\Lambda(x_0, u_0)$ 是非空的, 考虑由 (4.46) 定义的问题 (DL_d) 的最优解集 $\mathcal{S}(\text{DL}_d)$. (DL_d) 是问题 (PL_d) 的对偶, 在方向正则性条件之下, $\mathcal{S}(\text{DL}_d)$ 是 $\Lambda(x_0, u_0)$ 的非空的弱 * 紧致子集 (见命题 4.21).

我们还用下述的二阶条件: 存在常数 $\beta > 0$ 与 $\eta > 0$ 满足

$$\sup_{\lambda\in\mathcal{S}(\text{DL}_d)} D^2_{xx}L(x_0, \lambda, u_0)(h, h) \geqslant \beta\|h^2\|, \quad \forall h \in C_\eta(x_0) \tag{4.138}$$

与条件

$$\sup_{\lambda\in\mathcal{S}(\text{DL}_d)} \{D^2_{xx}L(x_0, \lambda, u_0)(h, h) - \sigma(\lambda, \mathcal{A}(h))\} > 0, \quad \forall h \in C(x_0) \setminus \{0\}, \tag{4.139}$$

其中, 对每一 $h \in C(x_0) \setminus \{0\}$,

$$\mathcal{A}(h) := \mathcal{A}_{K,M}(y_0, \delta) \tag{4.140}$$

是在点 $y_0 := G(x_0, u_0)$ 处沿方向 $\delta := D_x G(x_0, u_0)h$ 相对于映射 $M := DG(x_0, u_0) : X \times U \to Y$ 的集 K 的二阶近似集. 注意到, 这些条件显式地依赖于方向 d, 且较由 (3.136) 与 (3.155) 给出的相应的二阶条件要强, 因为 $\mathcal{S}(\mathrm{DL}_d)$ 是 $\Lambda(x_0, u_0)$ 的子集.

在本节余下部分, 除非额外说明, 考虑在参数空间 U 中沿方向 d 的下述形式的扰动

$$u(t) := u_0 + td + O(t^2), \quad t \geqslant 0. \tag{4.141}$$

注意到, 上述曲线 $u(t)$ 比 (4.133) 中的相应曲线满足更严格的条件.

定理 4.55 设 $\bar{x}(t)$ 是 $(\mathrm{P}_{u(t)})$ 的 $O(t^2)$ 最优解, 当 $t \downarrow 0$ 时收敛于 $x_0 \in \Phi(u_0)$, 且设

(i) Lagrange 乘子集合 $\Lambda(x_0, u_0)$ 是非空的, 方向正则性条件在 x_0 处沿方向 $d \in U$ 成立.

(ii) 或者 (a) 二阶充分性条件 (4.138) 成立, 或者 (b) 空间 X 是有限维的, 二阶充分条件 (4.139) 成立.

(iii) 线性化问题 (PL_d) 有最优解 $\hat{h} = \hat{h}(d)$, 满足对 $t \geqslant 0$,

$$\mathrm{dist}(G(x_0, u_0) + tDG(x_0, u_0)(\hat{h}, d), K) = O(t^2). \tag{4.142}$$

则

$$\|\bar{x}(t) - x_0\| = O(t). \tag{4.143}$$

注 4.56 上述定理的假设 (ii)(b) 类似于定理 3.83 中的假设, 除了在条件 (4.139) 中用的是较小的集合 $\mathcal{S}(\mathrm{DL}_d)$ 而不是 $\Lambda^g(x_0, u_0)$, 因而是较强的形式. 假设 (iii) 包含两部分, 即问题 (PL_d) 的最优解 $\hat{h}$ 的存在性与估计式 (4.142). 当然, 若二阶切集 $T_K^2(y_0, \hat{z})$ 是非空的, 其中 $\hat{z} := DG(x_0, u_0)(\hat{h}, d)$, 则估计式 (4.142) 成立. 注意, 因为 $\hat{h}$ 必满足 (PL_d) 的约束, 有 $\hat{z} \in T_K(y_0)$. 由定义 3.85, 若集 K 在 x_0 处是二阶正则的, 则对任何 $z \in T_K(y_0)$, 二阶切集 $T_K^2(y_0, z)$ 是非空的. 结果, 由 K 的二阶正则性可推出 (4.142).

定理 4.55 的证明基于最优值函数的下述上方估计.

引理 4.57 设 $x_0 \in \mathcal{S}(u_0)$, 即 x_0 是非扰动问题的最优解, 且设定理 4.55 中的假设 (i) 与 (iii) 成立. 则

$$v(u(t)) \leqslant v(u_0) + t\ \mathrm{val}(\mathrm{PL}_d) + O(t^2). \tag{4.144}$$

证明 令 $\hat{h}$ 满足 (4.142). 由方向正则性条件、引理 4.10 的估计 (4.17) 及假设 (4.142), 可推出存在 $\hat{x}(t) \in \Phi(u(t))$ 满足 $\|x_0 + t\hat{h} - \hat{x}(t)\| = O(t^2)$. 结果

$$v(u(t)) \leqslant f(\hat{x}(t), u(t)) = f(x_0, u_0) + tDf(x_0, u_0)(\hat{h}, d) + O(t^2). \tag{4.145}$$

因为 $v(u_0) = f(x_0, u_0), \hat{h}$ 是 (PL_d) 的最优解, 即得到不等式 (4.144). □

定理 4.55 的证明 此证明与定理 4.54 的证明采用相同的步骤. 若 (4.143) 不真, 则存在序列 $t_n \downarrow 0$, 置

$$x_n := \bar{x}(t_n), \quad u_n := u(t_n), \quad \tau_n := \|x_n - x_0\|, \quad h_n := \tau_n^{-1}(x_n - x_0),$$

有 x_n 是 (P_{u_n}) 的 $O(t_n^2)$ 最优解, 但 $\tau_n/t_n \to \infty$, 对任意 $\eta > 0$, 当 n 充分大时 $h_n \in C_\eta(x_0)$.

现在设二阶充分性条件 (4.138) 成立. 令 $\lambda \in \mathcal{S}(\mathrm{DL}_d)$. 因为 $\langle \lambda, G(x_n, u_n) - G(x_0, u_0)\rangle \leqslant 0, D_xL(x_0, \lambda, u_0) = 0$ 与 $t_n = o(\tau_n)$, 有

$$\begin{aligned} f(x_n, u_n) - f(x_0, u_0) \geqslant & L(x_n, \lambda, u_n) - L(x_0, \lambda, u_0) \\ = & t_n D_u L(x_0, \lambda, u_0) d \\ & + \frac{1}{2}\tau_n^2 D_{xx}^2 L(x_0, \lambda, u_0)(h_n, h_n) + o(\tau_n^2). \end{aligned} \tag{4.146}$$

因为 $\lambda \in \mathcal{S}(\mathrm{DL}_d)$, 所以 $D_uL(x_0, \lambda, u_0)d = \mathrm{val}(\mathrm{DL}_d)$, 由方向正则性有 $\mathrm{val}(\mathrm{PL}_d) = \mathrm{val}(\mathrm{DL}_d)$. 在有界集 $\mathcal{S}(\mathrm{DL}_d)$ 上极大化上式的右端, 由 (4.138) 得

$$f(x_n, u_n) - f(x_0, u_0) \geqslant t_n \mathrm{val}(\mathrm{PL}_d) + \frac{1}{2}\beta\tau_n^2 + o(\tau_n^2). \tag{4.147}$$

另一方面, 由于 x_n 是 $O(t_n^2)$ 最优解, 由 (4.144) 得

$$f(x_n, u_n) - f(x_0, u_0) \leqslant t_n \mathrm{val}(\mathrm{PL}_d) + O(t_n^2). \tag{4.148}$$

由于 $t_n = o(\tau_n)$, 不等式 (4.147) 与 (4.148) 是相互矛盾的. 这完成了二阶充分条件 (4.138) 情况下的证明. 现在设假设 (ii) 中充分条件的第二部分是成立的. 此种情况的下方估计的证明类似于定理 3.83 的证明中相应的推导. 我们讨论主要步骤. 因为空间 X 是有限维的, 如有必要可取子列, 设存在 $h \in C(x_0)$ 满足 $h \neq 0, x_n = x_0 + \tau_n h + \frac{1}{2}\tau_n^2 w_n$, 满足当 $\tau_n \downarrow 0$ 时 $\tau_n w_n \to 0$. 用 G 的二阶 Taylor 展式, 因为 $t_n = o(\tau_n)$, 可表示

$$\begin{aligned} G(x_n, u_n) = & y_0 + \tau_n D_x G(x_0, u_0) h + \frac{1}{2}\tau_n^2 D_x G(x_0, u_0) w_n \\ & + t_n D_u G(x_0, u_0) d + \frac{1}{2}\tau_n^2 D_{xx}^2 G(x_0, u_0)(h, h) + o(\tau_n^2). \end{aligned} \tag{4.149}$$

因为 $\mathcal{A}(h)$ 是集合 K 在 $G(x_0,u_0)$ 沿方向 $D_xG(x_0,u_0)h$ 关于映射 $DG(x_0,u_0)$: $X\times U\mapsto Y$ 的一个二阶近似集合, $t_n=o(\tau_n)$, 有

$$2t_nD_uG(x_0,u_0)d+\tau_n^2[D_xG(x_0,u_0)w_n+D_{xx}^2G(x_0,u_0)(h,h)]\in\tau_n^2\mathcal{A}(h)+o(\tau_n^2)B_Y. \tag{4.150}$$

令 $\lambda\in\mathcal{S}(\mathrm{DL}_d)$ 满足

$$D_{xx}^2L(x_0,\lambda,u_0)(h,h)-\sigma(\lambda,\mathcal{A}(h))=\kappa>0. \tag{4.151}$$

由 (4.150) 有

$$\begin{aligned}&\langle\lambda,\ 2t_nD_uG(x_0,u_0)d+\tau_n^2[D_xG(x_0,u_0)w_n+D_{xx}^2G(x_0,u_0)(h,h)]\rangle\\ \leqslant&\tau_n^2\sigma(\lambda,\mathcal{A}(h))+o(\tau_n^2).\end{aligned} \tag{4.152}$$

用二阶展式

$$\begin{aligned}f(x_n,u_n)=&f(x_0,u_0)+\tau_nD_xf(x_0,u_0)h+t_nD_uf(x_0,u_0)d\\ &+\frac{1}{2}\tau_n^2[D_xf(x_0,u_0)w_n+D_{xx}^2f(x_0,u_0)(h,h)]+o(\tau_n^2).\end{aligned} \tag{4.153}$$

连同 (4.151) 与 (4.152), 由于 $D_xf(x_0,u_0)h=0$(由 $h\in C(x_0)$ 得到), 得到

$$\begin{aligned}f(x_n,u_n)-f(x_0,u_0)\geqslant&t_nD_uf(x_0,u_0)d+\frac{1}{2}\tau_n^2[D_xf(x_0,u_0)w_n+D_{xx}^2f(x_0,u_0)(h,h)]\\ &+\frac{1}{2}\langle\lambda,\ 2t_nD_uG(x_0,u_0)d\\ &+\tau_n^2D_{xx}^2[D_xG(x_0,u_0)w_n+G(x_0,u_0)(h,h)]\rangle\\ &-\frac{1}{2}\tau_n^2\sigma(\lambda,\ \mathcal{A}(h))+o(\tau_n^2)\\ =&t_n\mathrm{val}(\mathrm{DL}_d)+\frac{1}{2}\tau_n^2D_{xx}^2L(x_0,\lambda,u_0)(h,h)\\ &-\frac{1}{2}\tau_n^2\sigma(\lambda,\ \mathcal{A}(h))+o(\tau_n^2)\\ =&t_n\mathrm{val}(\mathrm{DL}_d)+\frac{1}{2}\kappa\tau_n^2+o(\tau_n^2).\end{aligned}$$

因为 $\mathrm{val}(\mathrm{DL}_d)=\mathrm{val}(\mathrm{PL}_d)$, $t_n=o(\tau_n)$, 最后一不等式与 (4.148) 矛盾, 这样完成了证明. □

注 4.58 很清楚, 由上面的证明可得, $\mathcal{A}(h)$ 是关于映射 $DG(x_0,u_0)$ 的一个二阶近似集这一假设可松弛到映射

$$M(h,t):=D_xG(x_0,u_0)h+tD_uG(x_0,u_0)d.$$

在 K 的二阶正则性之假设下, 可取

$$\mathcal{A}(h) := T_K^2(G(x_0,u_0), D_xG(x_0,u_0)h)$$

作为一个二阶近似集合. 注 4.92 将说明对 $\mathcal{A}(h)$ 的这一选取, 二阶条件 (4.139) 对于 $\bar{x}(t)$ 的 Lipschitz 稳定性是 "几乎必要的".

注 4.59 上述证明中用到的定理 4.55 中的假设 (iii) 仅是为了建立最优值函数方向可微性质的上界估计 (4.144). 这一上界估计要比可表达为下述形式的上界估计 (4.51) 即

$$v(u(t)) \leqslant v(u_0) + t\ \mathrm{val}(\mathrm{PL}_d) + o(t)$$

强. 假设 (iii) 包含两部分, 即问题 (PL_d) 的最优解 $\hat{h}$ 的存在性与估计式 (4.142). 如指出的那样, 估计式 (4.142) 可由集合 K 的二阶正则性得到. 若 K 是多面体, 即在等式约束与有限多个不等式约束的情形, 假设 (iii) 总是成立的.

推出假设 (iii) 的另一个条件是严格约束规范 (4.119). 即相对于集合 $K_0 := \{y \in K | \langle\lambda,\ y - y_0\rangle = 0\}$ 的 Robinson 约束规范成立, 因此可以写为下述形式

$$D_xG(x_0,u_0)X - \mathcal{R}_{K_0}(y_0) = Y.$$

因此, 对任何 $d \in U$, 存在 $\hat{h} \in X$ 满足 $DG(x_0,u_0)(\hat{h},d) \in \mathcal{R}_{K_0}(y_0)$, 因此有 $DG(x_0, u_0)(\hat{h},d) \in T_K(y_0)$ 且 $\langle\lambda_0, DG(x_0,u_0)(\hat{h},d)\rangle = 0$. 于是 $\hat{h}$ 是 (PL_d) 的最优解 (见条件 (4.47)). 进一步, 存在某一 $t \geqslant 0$ 满足, $tDG(x_0,u_0)(\hat{h},d) \in K_0 - y_0$, 因此, 对充分小的 $t > 0$, (4.142) 左端的距离是零.

下述结果表明, 若空间 X 是自反的, 则 (PL_d) 最优解的存在性是近似解的 Lipschitz 稳定性的必要条件.

定理 4.60 *设方向正则性在点 $x_0 \in \mathcal{S}(u_0)$ 处沿方向 d 成立, 令 $u(t) = u_0 + td + o(t)$, 令 $\bar{x}(t)$ 是 ($\mathrm{P}_{u(t)}$) 的 $o(t)$ 最优解, 当 $t \downarrow 0$ 时收敛到 x_0, 则 $t^{-1}(\bar{x}(t) - x_0)$ 的任何弱聚点是问题 (PL_d) 的最优解. 相反地, 若 $\bar{h}$ 是 (PL_d) 的最优解且 $v'(u_0,d) = \mathrm{val}(\mathrm{PL}_d)$, 则存在 ($\mathrm{P}_{u(t)}$) 的 $o(t)$ 最优解 $\hat{x}(t)$ 满足 $\hat{x}(t) = x_0 + t\bar{h} + o(t)$.*

证明 设 $\bar{h}$ 是 $t^{-1}(\bar{x}(t) - x_0)$ 的弱聚点. 则存在序列 $t_n \downarrow 0$ 满足 $h_n := t_n^{-1}(x_n - x_0) \xrightarrow{w} \bar{h}$, 其中 $x_n := \bar{x}(t_n)$. 先证 $\bar{h}$ 满足问题 (PL_d) 的约束. 令 $u_n := u(t_n)$, 因为

$$G(x_n,u_n) - G(x_0,u_0) = t_nDG(x_0,u_0)(h_n,d) + o(t_n),$$

所以

$$t_n^{-1}(G(x_n,u_n) - G(x_0,u_0)) \xrightarrow{w} DG(x_0,u_0)(\bar{h},d). \tag{4.154}$$

因为 $G(x_n,u_n) \in K$, 所以 $G(x_n,u_n) - G(x_0,u_0) \in T_K(G(x_0,u_0)), T_K(G(x_0,u_0))$ 是弱闭的, 得到 $DG(x_0,u_0)(\bar{h},d) \in T_K(y_0)$.

证 $\bar{h}$ 的最优性. 不失一般性, 可设 $f(x_0,u_0)=0$, 即 $v(u_0)=0$, 则类似于 (4.154) 有

$$t_n^{-1}f(x_n,u_n)\to Df(x_0,u_0)(\bar{h},d),$$

从而

$$v(u_n)=f(x_n,u_n)+o(t_n)=t_nDf(x_0,u_0)(\bar{h},d)+o(t_n). \tag{4.155}$$

进一步, 由命题 4.22, 有 $v(u_n)\leqslant t_n\mathrm{val}(\mathrm{PL}_d)$, 这与 (4.155) 相连联系可得

$$Df(x_0,u_0)(\bar{h},d)\leqslant \mathrm{val}(\mathrm{PL}_d),$$

这表明 $\bar{h}$ 是 (PL_d) 的最优解.

相反地, 令 $\bar{h}$ 是 (PL_d) 的最优解, 则由定理 4.9,

$$\hat{x}(t):=x_0+t\bar{h}+r(t)\in\Phi(u(t))$$

对某一 $r(t)=o(t)$ 成立. 因为 $\mathrm{val}(\mathrm{PL}_d)=v'(u_0,d)$, 可得

$$f(\hat{x}(t),u(t))=tDf(x_0,u_0)(\bar{h},d)+o(t)=v(u(t))+o(t).$$

这表明 $\hat{x}(t)$ 是 $(\mathrm{P}_{u(t)})$ 的 $o(t)$ 最优解. □

注 4.61 若空间 X 自反的 (尤其是有限维的) 且存在 $(\mathrm{P}_{u(t)})$ 的 Lipschitz 稳定的 $o(t)$ 最优解 $\bar{x}(t)$, 则由紧致性, 可得序列 $t_n\downarrow 0$ 满足 $t_n^{-1}(\bar{x}(t_n)-x_0)$ 有弱极限. 由上述定理知, 这一弱极限是 (PL_d) 的最优解. 因此, 在自反空间 X 的情况, (PL_d) 最优解的存在性是 $(\mathrm{P}_{u(t)})$ 的解的 Lipschitz 稳定性的必要条件. 在例 4.54 中, 相应的线性化问题不具有最优解. 这解释了此例中为什么最优解不是 Lipschitz 稳定的.

注 4.62 由定理 4.25, 若 $\mathcal{S}(u_0)=\{x_0\}$, 方向正则性条件成立, 且 $\bar{x}(t)$ 是 Lipschitz 稳定的, 则 $v'(u_0,d)=\mathrm{val}(\mathrm{PL}_d)$. 上述定理的第二个结果表明, (PL_d) 的每一最优解对应 $(\mathrm{P}_{u(t)})$ 的 $o(t)$ 最优可行路径. 这说明, 若还有 X 是有限维空间, $o(t)$ 最优解的方向导数集合与线性化问题 (PL_d) 的最优解集 $\mathcal{S}(\mathrm{PL}_d)$ 是重合的. 然而, 很典型的是, 最优解集 $\mathcal{S}(\mathrm{PL}_d)$ 太大了, 以至于对最优解的方向导数的性质不能给出有用的信息. 例如, 对无约束问题, 有 $\mathcal{S}(\mathrm{PL}_d)=X$. 后面将会看到通过分析最优值函数的二阶方向导数, 如何计算最优解的方向导数.

下面的例子表明, 定理 4.55 中假设 (iii) 的条件 (4.142) 对于最优解的 Lipschitz 稳定性是本质性的.

例 4.63 考虑集合 $K\subset\mathbb{R}^3$,

$$K:=\{x=(x_1,x_2,x_3):x_3\geqslant[(x_1^2+x_2^2)^{1/2}-x_1]^{1+\alpha}\},$$

其中 $\alpha \in (0,1)$, 映射 $G(x_1,x_2,x_3,u) := (x_1,u,x_3)$ 由 $\mathbb{R}^3 \times \mathbb{R}$ 到 $\mathbb{R}^3$,

$$\min_{x \in \mathbb{R}^3} \ x_1^2 + x_2^2 + x_3 \quad \text{s. t.} \quad G(x,u) \in K \tag{4.156}$$

是凸问题, 对 $u_0 = 0$, 它有唯一最优解 $x_0 = (0,0,0)$. 对于 $u > 0$, 约束 $G(x,u) \in K$ 可写为下述形式

$$x_3 \geqslant [(x_1^2 + u^2)^{1/2} - x_1]^{1+\alpha},$$

因此, 此优化问题等价于

$$\min_{x_1} x_1^2 + [(x_1^2 + u^2)^{1/2} - x_1]^{1+\alpha}. \tag{4.157}$$

对于 $u > 0$, 上述问题有满足条件 $x_1 > 0$ 与下述条件的最优解 $\bar{x}_1(u)$ (通过置上述函数的导数为零)

$$2x_1 = (1+\alpha)[(x_1^2 + u^2)^{1/2} - x_1]^{\alpha}[1 - x_1(x_1^2 + u^2)^{-1/2}]. \tag{4.158}$$

由 (4.158) 可得, 当 $u \to 0$ 时, $\bar{x}_1(u)$ 以 $u^{(2+2\alpha)/(3+\alpha)}$ 的速率收敛到零, 因而不是 Lipschitz 稳定的.

另一方面, 在此例中, 除了 (4.142) 外, 定理 4.55 的其他条件均是满足的. 事实上, 不难验证 Slater 条件成立, 因而 Robinson 约束规范成立. 进一步, K 在零点处的切锥是 $\{(x_1,x_2,x_3) : x_3 \geqslant 0\}$, 因而线性化问题有最优解. 还有, 在 $x_0 = 0$ 处的临界锥是 $C(x_0) = \{(x_1,x_2,x_3) : x_2 = 0\}$, 二阶充分条件

$$D_{xx}^2 f(x_0,h_0)(h,h) > 0, \quad \forall h \in C(x_0) \setminus \{0\}$$

成立 (注意到, 在此例中, 约束是线性的, 对任何 Lagrange 乘子均有目标函数的 Hesse 阵与 Lagrange 函数的 Hesse 阵是相同的).

注 4.64 定理 4.55 中的假设不能保证沿形式 $u(t) := u_0 + td + o(t)$ 的路径的最优解的 Lipschitz 稳定性. 为了看到这一点, 先设 $d = 0$. 此种情况 $\mathcal{S}(\mathrm{DL}_d) = \Lambda(x_0,u_0)$, 方向正则性条件变为 Robinson 约束规范, 定理 4.55 的假设 (iii) 是自动成立的. 则在 Robinson 约束规范与二阶增长条件之下, 定理 4.55 的估计 (4.143) 表明, 对 $O(t^2)$ 阶的 $\|u(t) - u_0\|$, 距离 $\|\bar{x}(t) - x_0\|$ 是 $O(t)$ 阶的. 即 $\bar{x}(t)$ 是 $\frac{1}{2}$Hölder 稳定的, 这是命题 4.41 建立的结果. 例 4.32 表明, 在 Robinson 约束规范与二阶增长条件之下, 这一估计不可能被改善. 所以为了得到沿形式 $u(t) := u_0 + td + o(t)$ 的路径的 Lipschitz 稳定性, 需要提出定理 4.55 的正则性假设的一些 "一致的" 形式.

下述定理给出对定理 4.55 的参数序列 $u_n := u_0 + t_n d + o(t_n)$ 的这样一个一致形式. 对相应的方向序列 $d_n \longrightarrow d$, 考虑线性化问题序列 (PL_{d_n}) 以及它们的对偶 (DL_{d_n}).

定理 4.65 令 $u_n := u_0 + t_n d_n, d_n \longrightarrow d, t_n \downarrow 0, x_n$ 是收敛到 $x_0 \in \mathcal{S}(u_0)$ 的问题 (P_{u_n}) 的 $O(t_n^2)$ 最优解. 设

(i) Lagrange 乘子集合 $\Lambda(x_0, u_0)$ 非空, 且在 x_0 处沿方向 $d \in U$ 的方向正则性成立.

(ii) 对每一 d_n, (DL_{d_n}) 的最优解集 $\mathcal{S}(\mathrm{DL}_{d_n})$, 满足 $\mathcal{S}(\mathrm{DL}_d)$ 替换为 $\mathcal{S}(\mathrm{DL}_{d_n})$ 的二阶条件 (4.138) 成立, 常数 β 与 $n \in N$ 无关.

(iii) 线性化问题 (PL_{d_n}) 具有最优解 $\hat{h}_n$ 满足

$$\operatorname{dist}(G(x_0, u_0) + t_n DG(x_0, u_0)(\hat{h}_n, d_n), K) = o(t_n^2), \tag{4.159}$$

且序列 $\{\hat{h}_n\}$ 是有界的, 则

$$\|x_n - x_0\| = O(t_n). \tag{4.160}$$

证明 上方近似

$$v(u_n) \leqslant v(u_0) + t_n \mathrm{val}(\mathrm{PL}_{d_n}) + O(t_n^2) \tag{4.161}$$

可以类似于引理 4.57 中不等式 (4.144) 的证明得到证明. 另外, 由于 $\mathcal{S}(\mathrm{DL}_{d_n})$ 是 (PL_{d_n}) 的最优值函数的次微分, 方向正则性可推出 $\mathcal{S}(\mathrm{DL}_{d_n})$ 是一致有界的, 则用定理 4.55 的证明相同的推证, 在数据的展式中用 d_n 代替 d, 即证得 (4.160). □

注 4.66 若替换上述定理中的假设 (ii), 设空间 X 是有限维的, 且下述的二阶条件 (4.139) 的 "一致性" 扩展成立, 则同样的结论亦是成立的. 对 $h \in C(x_0) \setminus \{0\}$, 对所有充分大的 n,

$$\sup_{\lambda \in \mathcal{S}(\mathrm{DL}_{d_n})} \{D_{xx}^2 L(x_0, \lambda, u_0)(h, h) - \sigma(\lambda, \mathcal{A}(h))\} > 0, \tag{4.162}$$

其中

$$\mathcal{A}(h) := \mathcal{A}_{K,M}(G(x_0, u_0), D_x G(x_0, u_0)h)$$

是 K 在点 $G(x_0, u_0)$ 处沿方向 $D_x G(x_0, u_0)h$ 关于映射 $M := DG(x_0, u_0) : X \times U \to Y$ 的一个二阶近似集合.

沿扰动方向序列 $\{d_n\}$ 的有效条件可能是很难验证的. 因此, 给出用极限方向 d 自身来表达的充分条件是非常有用处的. 粗略地说, 下述引理表明, 这样的条件可以通过在二阶条件中变上确界为下确界的运算得到.

引理 4.67 设

(i) X 是自反的 Banach 空间.

(ii) $x_0 \in X$ 是满足 $\Lambda(x_0, u_0)$ 非空的一点.

(iii) 方向正则性在 x_0 处沿方向 d 成立.

(iv) 映射 $h \mapsto D_{xx}^2 G(x_0, u_0)(h,h)$ 在 X 的弱拓扑和 Y 的强拓扑下是连续的.

(v) 对每一 $\lambda \in \mathcal{S}(\mathrm{DL}_d)$, Hessian $D_{xx}^2 L(x_0, \lambda, u_0)(\cdot,\cdot)$ 是 Legendre 形式.

(vi) 对所有的 $\lambda \in \mathcal{S}(\mathrm{DL}_d)$, 有

$$D_{xx}^2 L(x_0, \lambda, u_0)(h,h) > 0, \quad \forall h \in C(x_0) \setminus \{0\}. \tag{4.163}$$

则存在 $\beta > 0$ 满足对充分接近于 d 的所有 d',

$$\inf_{\lambda \in \mathcal{S}(\mathrm{DL}_{d'})} D_{xx}^2 L(x_0, \lambda, u_0)(h,h) \geqslant \beta \|h\|^2, \quad \forall h \in C.\text{①} \tag{4.164}$$

证明　若 (4.164) 不真, 则存在 $d_n \longrightarrow d, \lambda_n \in \mathcal{S}(\mathrm{DL}_{d_n}), h_n \in C(x_0)$ 满足 $\|h_n\| = 1$, 有

$$D_{xx}^2 L(x_0, \lambda_n, u_0)(h_n, h_n) \leqslant o(1).$$

因为 X 是自反空间, 若有必要, 可抽取一子列, 不妨设 h_n 弱收敛到某一临界方向 $\bar{h}$ 满足 $\|\bar{h}\| \leqslant 1$. 由 (iii) 序列 λ_n 是有界的, 因而它至少有一弱 * 极限点 $\overline{\lambda}$, 所以由 (iv) 得

$$\begin{aligned} D_{xx}^2 L(x_0, \overline{\lambda}, u_0)(\bar{h}, \bar{h}) &= \lim_n D_{xx}^2 L(x_0, \overline{\lambda}, u_0)(h_n, h_n). \\ &= \lim_n D_{xx}^2 L(x_0, \lambda_n, u_0)(h_n, h_n) \leqslant 0. \end{aligned} \tag{4.165}$$

由于 $\Lambda(x_0, u_0)$ 对于弱 * 拓扑是闭的, 且包含集合 $\mathcal{S}(\mathrm{DL}_{d_n})$, 可得 $\overline{\lambda} \in \Lambda(x_0, u_0)$. 并且对每一 $\lambda \in \Lambda(x_0, u_0)$, 均有

$$\begin{aligned} D_u L(x_0, \overline{\lambda}, u_0) d &= \lim_{n\to\infty} D_u L(x_0, \lambda_n, u_0) d_n \geqslant \lim_{n\to\infty} D_u L(x_0, \lambda, u_0) d_n \\ &= D_u L(x_0, \lambda, u_0) d. \end{aligned}$$

这表明 $\overline{\lambda} \in \mathcal{S}(\mathrm{DL}_d)$. 然而由 (vi), $\bar{h} \in C(x_0)$ 与 (4.165) 可推出 $\bar{h} = 0$. 这一事实, (v) 与 (4.165) 也推出 h_n 是强收敛到 $\bar{h}$ 的, 满足 $\bar{h} \neq 1$, 从而与 (4.165) 矛盾. □

注 4.68　定理 4.65 中的假设 (iii) 在非线性规划, 即 K 是多面体集的情形下总是成立的. 此种情况, (PL_{d_n}) 是线性规划问题, 在方向正则性成立的前提下, 它总有一最优解 $\hat{h}_n$, 因此对于充分接近于 d 的 d_n, (PL_{d_n}) 的最优值是有限的. 进一步, 可以把 $\mathcal{S}(\mathrm{PL}_d)$ 写成为下式②

---

① (4.164) 中的 C 应为 $C(x_0)$.

② 应为 $\{h \in X : DG(x_0, u_0)(h,d) \in T_K(G(x_0,u_0)), Df(x_0,u_0)(h,d) \leqslant \mathrm{val}(\mathrm{PL}_d)\}$, 见 (4.451).

$$\{h \in X : Dg(x_0, u_0)(h, d) \in T_K(G(x_0, u_0)), Dg(x_0, u_0)(h, d) \leqslant \mathrm{val}(\mathrm{PL}_d)\}.$$

应用 Hoffman 引理 (定理 2.200) 得, 存在 $\bar{h} \in \mathcal{S}(\mathrm{PL}_d)$, 只要 $d' \to d$, 就有 $\mathrm{dist}(\bar{h}, \mathcal{S}(\mathrm{PL}_{d'})) \to 0$. 从而可以设序列 $\{\hat{h}_n\}$ 是有界的. 因为 K 是多面体, 条件 (4.159) 亦是成立的.

注 4.69 由命题 4.50(ii) 可得, 在条件 (4.125) 成立时, 线性化问题 (PL_{d_n}) 具有有界的最优解序列的另外一种情况 (由定义 4.46 给出的严格约束规范可推出条件 (4.125)). 然而, 这种情况之下, 仍然要验证假设 (4.159).

最后作如下的观察. 令 $\bar{x}(u)$ 是收敛到 x_0 的 (P_u) 的最优解. S 是 U 的紧致子集. 设对每一 d, $u(t) := u_0 + td + o(t)$, 相应的最优解 $\bar{x}(u(t))$ 是 Lipschitz 稳定的. 则由紧致性推证容易验证 $\bar{x}(u)$ 在集合 S 生成的锥 $\mathcal{C} := \mathrm{cone}(S)$ 是一致 Lipschitz 稳定的, 即存在 $\varepsilon > 0$ 满足

$$\|\bar{x}(u_0 + d) - x_0\| \leqslant c\|d\|, \quad \forall\, d \in \mathcal{C} \cap (\varepsilon B_U). \tag{4.166}$$

4.6 通过一种简化方式的量化稳定性分析

在 3.4.4 节中我们已经讨论过一种情况, 其中集合 $K \subset Y$ 在点 $y_0 \in Y$ 的一个邻域 N 内可局部参数化为一个较简单的集合 $\mathcal{C} \subset Z$, 即存在一足够光滑的映射 $\Xi(y)$, 从 y_0 的邻域 N 到 Z 满足 $K \cap N = \Xi^{-1}(\mathcal{C})$ 且 $D\Xi(y_0)$ 是映上的 (见定义 3.135). 这一节用这种简化方式来研究相应的参数化最优化问题最优解的稳定性.

对非扰动问题 (P) 引入非退化性的概念, 证明在此非退化性条件下, 问题 (P_u) 可局部地参数化表示为新的问题, 它的可行集不依赖于 u. 即在此非退化情形之下, 最优解的 Lipschitz 稳定性问题可简化为研究固定可行集的参数化问题. 这一节, 设约束映射 $G : X \times U \to Y$ 是连续可微的, 且为简化起见, 设空间 X, Y 与 Z 是有限维的, 如 $X = \mathbb{R}^m$, $Y = \mathbb{R}^n$, $Z = \mathbb{R}^k$. 对非扰动问题, 用 $G(\cdot)$ 表示 $G(\cdot, u_0)$ 等.

4.6.1 非退化性与严格互补性

考虑非扰动问题

$$\text{(P)} \qquad \min_{x \in X} f(x) \qquad \text{s. t.} \qquad G(x) \in K \tag{4.167}$$

的可行点 $x_0 \in \Phi$. 非正规地说, 称 x_0 是映射 $G(x)$ 关于集合 K 的一个非退化点, 若在 Y 中可以构造局部坐标系统 $\xi_1(y), \cdots, \xi_k(y), k \leqslant n$, 满足在 $y_0 := G(x_0)$ 的附近, 集合 K 可以用这一坐标系统来描述, 且复合映射 $x \mapsto (\xi_1(G(x)), \cdots, \xi_k(G(x)))$ 于 x_0 处在定义 4.70 的意义下非退化.

设集合 K 在点 $y_0 \in K$ 处以定义 3.135 的意义, 被映射 $\Xi(y) = (\xi_1(y), \cdots, \xi_k(y))C^1$ 简约到集合 $\mathcal{C} \subset \mathbb{R}^k$. 即 $D\Xi(y_0) : Y \to \mathbb{R}^k$ 是映上的, $K \cap N = \{y \in N :$

$\Xi(y) \in \mathcal{C}\}$ 且 $\Xi(y_0) = 0$. 考虑集合

$$W := \{y \in N : \xi_i(y) = 0, i = 1, \cdots, k\}. \tag{4.168}$$

因为 $\Xi(y_0) = 0$, 有 $y_0 \in W$ 且 $W \subset K$. 进一步, 由于 $D\Xi(y_0)$ 是映上的, 即梯度 $\nabla\xi_1(y_0), \cdots, \nabla\xi_k(y_0)$ 是线性无关的, 则 W 在 y_0 附近为 $n-k$ 维的光滑流形, 其相应的切空间 $T_W(y_0)$ 为

$$T_W(y_0) = \ker[D\Xi(y_0)], \tag{4.169}$$

其中 $\ker\,[D\Xi(y_0)] := \{h \in Y : D\Xi(y_0)h = 0\}$.

定义 4.70　设 K 在 $y_0 \in K$ 被映射 $\Xi(y) = (\xi_1(y), \cdots, \xi_n(y))C^1$ 简约到集合 $\mathcal{C} \subset \mathbb{R}^k$. 称 x_0 是映射 $G(x)$ 关于集合 K 与映射 Ξ 的非退化点, 若

$$DG(x_0)X + \ker[D\Xi(y_0)] = Y. \tag{4.170}$$

注意到, 由 (4.169), 条件 (4.170) 可等价地表示为

$$DG(x_0)X + T_W(y_0) = Y. \tag{4.171}$$

上述条件意味着 G 在点 x_0 处与光滑流形 W 横截性地相交. 还注意到

$$D\Xi(y_0)(DG(x_0)X + \ker[D\Xi(y_0)]) = D\Xi(y_0)(DG(x_0)X),$$

并由链式法则, 可得 $D(\Xi \circ G)) = D\Xi \circ DG$. 所以, 条件 $D\Xi(y_0)Y = \mathbb{R}^k$ 与 (4.170) 同时成立当且仅当 $D(\Xi \circ G)(x_0) : X \to \mathbb{R}^k$ 是映上的, 即点 x_0 是非退化的充要条件是 K 在 y_0 的附近可局部地由坐标系统 $\xi_i, i = 1, \cdots, k$ 来描述, 且梯度 $\nabla\gamma_1(x_0), \cdots, \nabla\gamma_k(x_0)$ 是线性无关的, 其中 $\gamma_i(x) := \xi_i(G(x))$. 则问题 (P) 的可行集 Φ 在点 x_0 的附近可由约束 $\mathcal{G}(x) \in \mathcal{C}$ 定义, 其中 $\mathcal{G}(x) := \Xi(G(x))$, 且集合 $\mathcal{G}^{-1}(0) \subset \Phi$ 是点 x_0 邻域内的光滑流形.

需要指出的是, 非退化性的上述定义涉及映射 G 与集合 K, 也依赖于局部坐标系统 $\xi_1(y), \cdots, \xi_k(y)$ (定义的简约映射) 特定的选取. 下面的例子将说明这一点.

例 4.71　考虑集合 $K := \{(y_1, y_2) : y_1 \geqslant 0\} \subset \mathbb{R}^2$, 从 $\mathbb{R}$ 到 $\mathbb{R}^2$ 的映射 $G(x) := (x, 0)$. 显然, $\Phi := \{x : x \geqslant 0\}$. 考虑 $\mathbb{R}^2$ 中的标准坐标系统, 即 $\xi_1(y_1, y_2) := y_1$, $\xi_2(y_1, y_2) := y_2$. 则 $\Xi \circ G$ 是线性映射, $\Xi \circ G : x \to (x, 0)$, 因此对任何 $x_0 \in \mathbb{R}$, $D(\Xi \circ G)(x_0) = \Xi \circ G$ 不是映上的. 另一方面, 集合 K 由约束 $y_1 \geqslant 0$ 定义, 因此可取 $\xi_1(y_1, y_2) := y_1, k = 1$. 对这样的坐标选取, $\Xi \circ G(x) = x$, 因而非退化性成立. 当然, 在前一种表示中, 坐标 y_2 是多余的, 一维表示更好些.

因为在 y_0 的邻域中, 由 (4.168) 定义的集合 W 是 K 的子集, 有 $T_W(y_0) \subset T_K(y_0)$. 进一步, 如命题 4.73 所示, 假若这一简约是点的, $T_W(y_0)$ 是切锥 $T_K(y_0)$ 的线空间, 即包含在 $T_K(y_0)$ 中的最大的线性空间 (简约被称为是点的, 若切锥 $T_C(\Xi(y_0))$ 是一点锥). 因此, 若简约是点的, 则点 x_0 是非退化的充分必要条件是

$$DG(x_0)X + \text{lin}(T_K(y_0)) = Y. \tag{4.172}$$

上述方法不依赖于简约映射Ξ. 因此对点简约, 条件 (4.172) 可用作非退化性的定义.

注 4.72 设问题 (P) 的可行集定义为形式 $x \in K_1$, $G(x) \in K_2$, 其中 $K_1 \subset X$, $K_2 \subset Y$ 是闭凸集. 这些约束可写为 $\overline{G}(x) \in K$, 其中 $K := K_1 \times K_2$, $\overline{G}(x) := (x, G(x))$. 令 x_0 是 (P) 的可行点, 设 K_1 与 K_2 在 x_0 与 $y_0 := G(x_0)$ 处分别是 C^1 简约的, 这两个简约是点的. 则由 (4.172), 点 x_0 关于 $\overline{G}(x)$ 与 K 是非退化的, 当且仅当对任何 $(x, y) \in X \times Y$, 存在 $h \in X$ 满足

$$\begin{pmatrix} x \\ y \end{pmatrix} \in \begin{pmatrix} h \\ DG(x_0)h \end{pmatrix} + \begin{pmatrix} \text{lin}(T_{K_1}(x_0)) \\ \text{lin}(T_{K_2}(y_0)) \end{pmatrix}.$$

上述条件成立当且仅当

$$DG(x_0)[x + \text{lin}(T_{K_1}(x_0))] + \text{lin}(T_{K_2}(y_0)) = Y,$$

它等价于

$$DG(x_0)[\text{lin}(T_{K_1}(x_0))] + \text{lin}(T_{K_2}(y_0)) = Y. \tag{4.173}$$

因此得到, 点 x_0 是非退化的充分必要条件是 (4.173) 成立.

命题 4.73 令 K 是 $\mathbb{R}^n$ 的闭凸子集, $y_0 \in K$. 则

$$[\text{lin}(T_K(y_0))]^{\perp} = \text{Sp }(N_K(y_0)), \tag{4.174}$$

若 $\lambda \in \text{ri}(N_K(y_0))$, 则

$$\text{lin}(T_K(y_0)) = T_K(y_0) \cap \ker\lambda. \tag{4.175}$$

进一步, 若 K 在 y_0 处 C^1 简约为集合 C 且简约是点的, 则

$$\text{lin}(T_K(y_0)) = T_W(y_0). \tag{4.176}$$

证明 向量 w 属于空间 Sp $(N_K(y_0))$ 的直交补空间当且仅当对所有的 $\lambda \in N_K(y)$, $\langle \pm w, \lambda \rangle \leqslant 0$. 这意味着 $\pm w$ 属于 $T_K(y_0)$. 所以, $\text{lin}(T_K(y_0)) = [\text{Sp }(N_K(y_0))]^{\perp}$, 因此, 可通过在此式两边取正交补得到 (4.174).

若 $\lambda \in N_K(y_0)$, $w \in \text{lin}(T_K(y_0))$, 则 $\langle w, \lambda \rangle \leqslant 0$ 且 $\langle -w, \lambda \rangle \leqslant 0$, 因此 $\langle w, \lambda \rangle = 0$. 于是 $\text{lin}(T_K(y_0)) \subset T_K(y_0) \cap \text{ker}\lambda$. 现在设 $\lambda \in \text{ri}(N_K(y_0))$, $w \in T_K(y_0) \cap \text{ker}\lambda$. 则对任何 $\mu \in \text{Sp}\ (N_K(y_0))$, 当 $t > 0$ 充分小时, $\lambda + t\mu$ 与 $\lambda - t\mu$ 属于 $N_K(y_0)$, 因此 $\langle w, \lambda + t\mu \rangle \leqslant 0$ 且 $\langle w, \lambda - t\mu \rangle \leqslant 0$. 这就得到 $\langle w, \mu \rangle = 0$, 因此 $w \in [\text{Sp}\ (N_K(y_0))]^{\perp}$. 公式 (4.175) 由 (4.174) 得到.

我们有 $T_W(y_0) \subset T_K(y_0)$. 因为 $D\Xi(y_0)(T_K(y_0)) = T_C(\Xi(y_0))$, 且 $T_C(\Xi(y_0))$ 是点的, $T_K(y_0)$ 的任何线性子空间被 $D\Xi(y_0)$ 映射到 $\{0\}$. 则由 (4.169) 得 $T_K(y_0)$ 的线空间包含在 $T_W(Y_0)$ 中. 结果, 切锥 $T_K(y_0)$ 的线空间与 $T_W(y_0)$ 是重合的, 证毕. □

若此时设 $K = \mathbb{R}^n_-$, 即 (P) 是可行集由有限个不等式约束定义的非线性规划问题, 则称严格互补条件在可行点 x_0 处成立, 若存在 Lagrange 乘子向量 $\lambda = (\lambda_1, \cdots, \lambda_n) \in \Lambda(x_0)$, 满足对所有的 $i \in I(x_0)$, 有 $\lambda_i > 0$, 其中 $I(x_0) := \{i | g_i(x_0), i = 1, \cdots, n\}$ 是在 x_0 处起作用约束的指标集. 有

$$N_K(y_0) = \{y \in \mathbb{R}^n : y_i \geqslant 0, i \in I(x_0), y_i = 0, i \notin I(x_0)\},$$

因此严格互补条件意味着 $\lambda \in \text{ri}(N_K(y_0))$. 我们用这一性质作为严格互补条件的一般性定义.

定义 4.74 *称严格互补条件在问题 (P) 的可行点 x_0 处成立, 若存在 Lagrange 乘子 $\lambda \in \Lambda(x_0)$ 满足 $\lambda \in \text{ri}(N_K(y_0))$.*

因为 $T_W(Y_0) \subset T_K(y_0)$, 由 (4.171) 可得

$$DG(x_0)X + T_K(Y_0) = Y, \tag{4.177}$$

这等价于 (因为 Y 是有限维的)Robinson 约束规范 (见命题 2.97), 即非退化条件比 Robinson 约束规范要强. 因此, 若非退化点 x_0 是 (P) 的局部最优解, 则存在 Lagrange 乘子 λ 满足相对应的一阶必要条件. 实际上, 非退化性条件可推出相应的 Lagrange 乘子的唯一性.

命题 4.75 *令 x_0 是 (P) 的局部最优解, 设集合 K 在点 $y_0 := G(x_0)$ 处 C^1 简约为凸闭集 C. 则下述结论成立:*

(i) 若点 x_0 是非退化的, 则 $\Lambda(x_0)$ 是单点集.

(ii) 若 $\Lambda(x_0)$ 是单点集, 严格互补条件成立, 且此简约是点的, 则点 x_0 是非退化的.

证明 因为非退化条件可推出 Robinson 约束规范, 则集合 $\Lambda(x_0)$ 是非空的. 令 $\lambda \in \Lambda(x_0)$. 因为 $\lambda \in N_K(y_0), T_W(y_0) \subset T_K(y_0)$, 有 λ 与 $T_W(y_0)$ 是直交的. 若 $\lambda' \in \Lambda(x_0)$ 是另一 Lagrange 乘子, 则 $\lambda - \lambda'$ 与 $DG(x_0)X$ 和 $T_W(y_0)$ 均是垂直的. 由 (4.171) 即得 $\lambda - \lambda'$ 与 Y 垂直, 从而有 $\lambda - \lambda' = 0$. 这证得 Lagrange 乘子的唯一性.

现在来证结论 (ii). 令 $\lambda \in \mathrm{ri}(N_K(y_0))$ 是 Lagrange 乘子. 因为此简约是点的, 由命题 4.73 得 $\mathrm{Sp}\ (N_K(y_0)) = T_W(y_0)^\perp$. 通过对 (4.171) 两边取正交补得, 它等价于

$$\ker[DG(x_0)^*] \cap [T_W(y_0)^\perp] = \{0\}. \tag{4.178}$$

假设点 x_0 不是非退化的. 则 (4.178) 不成立, 即 (4.178) 的左端含有一向量 $\mu \neq 0$. 因为 $\mu \in \ker[DG(x_0)^*]$, 则对任何 $t \in \mathbb{R}$, $D_x L(x_0, \lambda + t\mu) = 0$. 因 $\mu \in \mathrm{Sp}\ (N_K(y_0))$, $\lambda \in \mathrm{ri}(N_K(y_0))$, 对 $t > 0$ 充分小有 $\lambda + t\mu \in N_K(y_0)$, 因此 $\lambda + t\mu$ 是 Lagrange 乘子, 这与 λ 的唯一性矛盾. □

简约是点的这一假设在上述命题结论 (ii) 的证明中起本质性作用, 因为否则我们仅能得到 $T_W(y_0)$ 包含在 $T_K(y_0)$ 的线空间中.

注 4.76 设 x_0 是非退化点, λ_0 是相应的 (唯一的)Lagrange 乘子. 得到

$$T_W(y_0) \subset \mathrm{lin}(T_K(y_0)) \subset T_K(y_0) \cap \ker\lambda_0.$$

因此, 由 x_0 的非退化性可推出下述条件

$$DG(x_0)X + T_K(y_0) \cap \ker\lambda_0 = Y. \tag{4.179}$$

上述条件 (4.179) 恰好是条件 (4.125), 因此命题 4.50 的结论可用. 还注意到, 若 $\lambda_0 \in \mathrm{ri}(N_K(y_0))$, 即严格互补条件成立, 且简约是点的, 则由 (4.175) 与 (4.176), 有 $T_K(y_0) \cap \ker\lambda_0 = T_W(y_0)$.

现在讨论 3.4.4 节中的例子.

例 4.77 考虑例 3.138 的情况, 即设集合 K 在 y_0 附近由有限多个不等式约束 $\xi_1(y) \leqslant 0, \cdots, \xi_k(y) \leqslant 0$ 定义, 其中 $\xi_i(y)$, $i = 1, \cdots, k$ 是连续可微函数. 问题 (P) 的可行集 Φ 在点 x_0 附近由约束 $\gamma_i(x) \leqslant 0$, $i = 1, \cdots, k$ 定义, 其中 $\gamma_i(x) = \xi_i(G(x))$. 此种情况, 点 x_0 是非退化的充分必要条件是 $\nabla\xi_1(y_0), \cdots, \nabla\xi_k(y_0)$ 线性无关, 条件 (4.171) 成立, 且

$$T_W(y_0) = \{\eta : D\xi_i(y_0)\eta = 0,\ i = 1, \cdots, k\}^{①}.$$

等价地, x_0 是非退化的充分必要条件是 $\nabla\gamma_1(y_0), \cdots, \nabla\gamma_k(y_0)$ 是线性无关的. 严格互补条件意味着 $\lambda_i > 0$ 对所有的 $i \in I(y_0)$ 成立.

例 4.78 考虑例 3.139 中的情况, 即设集合 K 是多面体. 令 $L := \mathrm{lin}[T_K(y_0)]$ 是 $T_K(y_0)$ 的线空间, $L^\perp$ 是其正交补空间. 考虑 $L^\perp$ 中的一组基 $b_1, \cdots, b_k$, 考虑坐标系统 $\xi_i(y) := b_i^{\mathrm{T}}(y - y_0)$, $i = 1, \cdots, k$. 此坐标系统中, x_0 是非退化的当且仅当

$$DG(x_0)X + \mathrm{lin}[T_K(y_0)] = Y. \tag{4.180}$$

① 原著中为 $D\xi_i(y)\eta$.

尤其, 若 $K := \mathbb{R}^n_-$, 则 x_0 是非退化的充分必要条件是梯度向量 $\nabla g_i(x_0),\ i \in I(x_0)$ 是线性无关的, 其中 $I(x_0) := \{i : g_i(x_0) = 0,\ i = 1, \cdots, n\}$, 严格互补条件成立的充分必要条件是对 $i \in I(x_0)$ 有 $\lambda_i > 0$.

例 4.79 考虑半定规划的例子 (见例 3.140), 即令 $Y := S^p$ 是 $p \times p$ 对称矩阵构成的空间, $K := S^p_-$ 是 $p \times p$ 负半定对称矩阵构成的锥. 令 x_0 是 (P) 的可行点, 令 $A_0 := G(x_0) \in S^p_-$ 的秩是 $r < p$. 考虑简约映射 $\Xi(A) = U(A)^{\mathrm{T}} AU(A)$, 它由 A_0 的邻域 N 到 S^{p-r}, 此映射的构造见例 3.140. 回顾 $D\Xi(A_0)Z = E_0^{\mathrm{T}} ZE_0$, 其中 $E_0 = U(A_0)$ 是秩为 $p - r$ 的 $p \times (p - r)$ 阵, 满足 $A_0E_0 = 0$. 集合 $W := \{A \in N : \Xi(A) = 0\}$ 由秩为 r 的对称矩阵构成, 是光滑的流形, 且

$$T_W(A_0) = \{Z \in S^p : E_0^{\mathrm{T}} ZE_0 = 0\}. \tag{4.181}$$

在这样的简约之下得到可行点 x_0 是非退化的充分必要条件是

$$DG(x_0)X + T_W(A_0) = S^p. \tag{4.182}$$

条件 (4.182) 意味着 G 在 x_0 处与秩为 r 的对称矩阵构成的光滑流形横截地相交. 因为

$$T_{S^p_-}(A_0) = \{Z \in S^p : E_0^{\mathrm{T}} ZE_0 \preceq 0\}, \tag{4.183}$$

$$N_{S^p_-}(A_0) = S^p_+ \cap A_0^{\perp} = \{E_0QE_0^{\mathrm{T}} : Q \in S^{p-r}_+\}. \tag{4.184}$$

所以

$$\mathrm{ri}(N_{S^p_-}(A_0)) = \{Z = E_0QE_0^{\mathrm{T}} : Q \in S^{p-r}_+, Q \succ 0\}. \tag{4.185}$$

由于这一简约是点的, 有

$$T_W(A_0) = \mathrm{lin}(T_{S^p_-}(A_0)). \tag{4.186}$$

令 x_0 是可行点, $A_0 := G(x_0)$, 且 $\Omega \in \Lambda(x_0)$ 是相应的 Lagrange 乘子矩阵. 因为 $\Omega \in N_{S^p_-}(A_0)$, 由 (4.184) 得

$$\mathrm{rank}(\Omega) + \mathrm{rank}(G(x_0)) \leqslant p. \tag{4.187}$$

进一步, 由 (4.185) 可得, 严格互补条件在 x_0 处成立当且仅当存在 $\Omega \in \Lambda(x_0)$ 满足

$$\mathrm{rank}(\Omega) + \mathrm{rank}(G(x_0)) = p. \tag{4.188}$$

4.6.2 稳定性分析

现在回到参数化问题 (P_u), 证明在非退化的条件下, 问题 (P_u) 可局部地参数化表示为新问题, 它的可行集不依赖于参数向量 u.

定理 4.80 设 x_0 是 $G(\cdot,u_0)$ 关于 K 与 $\Xi(y)$ 的非退化点. 则存在 x_0 的邻域 $\mathcal{X}$, u_0 的邻域 $\mathcal{U}$, $0\in\mathbb{R}^m$ 的邻域 $\mathcal{Z}$ 及连续可微映射 $T:\mathcal{Z}\times\mathcal{U}\to\mathcal{X}$ 满足 $T(0,u_0)=x_0$, 且对每一 $u\in\mathcal{U}$, $T(\cdot,u)$ 是从 $\mathcal{Z}$ 到 $\mathcal{X}$ 上的微分同胚 (即 $T(\cdot,u):\mathcal{Z}\to\mathcal{X}$ 是一一对应的、映上的, 其逆也是可微的), 且

$$T((\mathcal{C}\times\mathbb{R}^{m-k})\cap\mathcal{Z},u)=\Phi(u)\cap\mathcal{X}. \tag{4.189}$$

证明 考虑定义于 (x_0,u_0) 的邻域上的映射 $H(x,u):=(\Xi\circ G)(x,u)$. 因为 x_0 是非退化的, 所以 $D_xH(x_0,u_0)$ 是映上的, 或等价地, 其转置由 $m\times k$ 阶 Jacobi 阵 $\nabla_xH(x_0,u_0)$ 给出, 秩是 k 阶的. 不妨设 $\nabla_xH(x_0,u_0)$ 的上 $k\times k$ 阶子阵是非奇异的. 考虑下述方程系统

$$z_1=h_1(x,u),\cdots,z_k=h_k(x,u),\ z_{k+1}=x_{k+1},\cdots,z_m=x_m, \tag{4.190}$$

其中 $h_i(x,u)$ 是映射 $H(x,u)$ 的坐标函数. 有此系统相对于 x 的 Jacobi 阵在 (x_0,u_0) 处是非奇异的, 则由隐函数定理, 对接近于 u_0 的 u, 上述方程可以局部地反求出 x. 即对接近于 u_0 的 u, 存在局部的微分同胚 $T(\cdot,u):\mathcal{Z}\to\mathcal{X}$ 满足 $z_i=h_i(T(z,u),u)$, $i=1,\cdots,k$, 且 $T(z,u)$ 是连续可微的. 因为局部地, $\Phi(u)$ 可由约束 $(h_1(x,u),\cdots,h_k(x,u))\in\mathcal{C}$ 来定义, 所以方程 (4.189) 成立. □

上述定理表明, 在 x_0 附近, 对 u_0 的邻域中的所有 u, 问题 (P_u) 可重新参数化为下述等价的问题

$$\min_{z\in\mathbb{R}^m} f(T(z,u),u) \qquad \text{s. t.} \quad (z_1,\cdots,z_k)\in\mathcal{C}, \tag{4.191}$$

其可行域不依赖于 u.

定理 4.80 所述的重新参数化过程与命题 4.32 可以推出当最优解集 $S_0:=\mathcal{S}(u_0)$ 不必是单点集时的下述稳定性结果. 函数 (映射)$g(\cdot)$ 称为 $C^{1,1}$ 的, 若它是连续可微的且 $Dg(\cdot)$ 是局部 Lipschitz 连续的. $C^{1,1}$ 函数类介于 C^2(二次连续可微) 函数类与 C^1(连续可微) 函数类之间.

定理 4.81 设

(i) $f(x,u)$ 与 $G(x,u)$ 是 $C^{1,1}$ 类的.

(ii) 非扰动问题有一非空的最优解集 $S_0:=\mathcal{S}(u_0)$.

(iii) (非扰动问题的) 二阶增长条件在 S_0 处成立.

(iv) 集合 S_0 中的每一点均是 $G(\cdot,u_0)$ 相对于 K 的非退化点, 相应的坐标函数 $\xi_i(y)$ 是 $C^{1,1}$ 的函数.

(v) 对 u_0 某一邻域中所有的 u, 最优解集 $\mathcal{S}(u)$ 是一致有界的.

则 $\mathcal{S}(u)$ 在 u_0 处是 Lipschizt 稳定的.

证明 考虑序列 $u_n \to u_0$, 令 $x_n \in \mathcal{S}(u_n)$. 只需证明

$$\operatorname{dist}(x_n, S_0) = O(\|u_n - u_0\|). \tag{4.192}$$

由于 $\mathcal{S}(u)$ 是一致有界的, 如有必要可取一子列, 设序列 $\{x_n\}$ 收敛到一点 x_0. 由连续性结论 (见 4.1 节), 有 $x_0 \in S_0$. 由假设有点 x_0 是非退化的, 由定理 4.80, 存在连续可微映射 $T(z,u)$, 满足在 x_0 的附近对 u_0 邻域中所有的 u, 问题 (P_u) 可重新参数化为问题 (4.191). 注意到, 对充分大的 n, 关系 $T(z_n, u_n) = x_n$ 定义了当 $u = u_n$ 时问题 (4.191) 的最优解序列 $z_n \to 0$. 令 Z_0 是限定到 0 的邻域的重新参数化问题 (4.191) 的最优解集合. 因为 $T(\cdot, u)$ 在 $z = 0$ 附近是 Lipschitz 连续的, 其 Lipschitz 常数对于 u_0 某一邻域中所有的 u 是一致有界的, 估计式 (4.192) 可由下式推出

$$\operatorname{dist}(z_n, Z_0) = O(\|u_n - u_0\|). \tag{4.193}$$

剩下要证 Lipschitz 稳定性结果 (4.193) 对重新参数化问题成立. 首先观察到映射 $T(\cdot, u_0)$ 与它的逆是连续可微的, 因此在 x_0 的附近是 Lipschitz 连续的, 则问题 (4.191) 对 $u = u_0$ 限制在 $z = 0$ 邻域的二阶增长条件由问题 (P) 相应的二阶增长条件推出. 因此, 由命题 4.32 的上界估计 (4.80), 只需要检验在 0 的一个邻域 $\mathcal{Z}$ 上的函数 $h(\cdot, u) := f(T(\cdot, u), u) - f(T(\cdot, u_0), u_0)$ 的 Lipschitz 常数是 $O(\|u - u_0\|)$ 阶的. 由中值定理, 这一 Lipschitz 常数可由 $\sup\limits_{z \in \mathcal{Z}} \|\nabla_z h(z, u)\|$ 给出. 因此, 由微分的链式法则, 由于 f 是 $C^{1,1}$ 函数, 若

$$\sup_{z \in \mathcal{Z}} \|\nabla_z T(z, u) - \nabla_z T(z, u_0)\| = O(\|u_n - u_0\|), \tag{4.194}$$

则 $h(\cdot, u)$ 在 $\mathcal{Z}$ 上的 Lipschitz 常数是 $O(\|u_n - u_0\|)$ 阶的. 由于坐标映射 Ξ 被设为是 $C^{1,1}$ 的, 复合映射 $\Xi \circ G$ 也是 $C^{1,1}$ 的. 隐函数定理可推出映射 $T(z, u)$ 亦是 $C^{1,1}$ 的. 证毕. □

上述定理给出的稳定性结果可推广到当 $\varepsilon = O(\|u - u_0\|^2)$ 时 (P_u) 的 ε 最优解集上.

令 $x_0 \in \mathcal{S}(u_0)$ 是 $G(\cdot, u_0)$ 的非退化点, $\bar{x}(u)$ 是 (P_u) 的最优解. 则对 u_0 邻域中的 u, 充分接近 x_0 的 $\bar{x}(u)$ 对应于唯一的 Lagrange 乘子 $\overline{\lambda}(u)$. 下面讨论 (P_u) 的局部最优解和相应的 Lagrange 乘子与下述简约问题的最优解和相应的 Lagrange 乘子之间的关系

$$(\mathcal{P}_u) \qquad \min_{x \in X} f(x, u) \quad \text{s. t.} \quad \mathcal{G}(x, u) \in \mathcal{C}, \tag{4.195}$$

其中 $\mathcal{G}(x, u) := \Xi(G(x, u))$. 由定义 4.70 的性质 (i), 对充分接近于 u_0 的 u, (P_u) 与 $(\mathcal{P}_u)$ 的可行域在 x_0 附近是重合的. 所以, 限定在 x_0 的邻域上, 对于充分接近于 u_0 的所有的 u, (P_u) 与 $(\mathcal{P}_u)$ 的最优解集是相同的.

令 $\bar{\mu} = \bar{\mu}(u)$ 是对应于 $\bar{x} = \bar{x}(u)$ 的 Lagrange 乘子向量, 满足一阶最优性条件

$$D_x\mathcal{L}(\bar{x}, \bar{\mu}, u) = 0, \quad \bar{\mu} \in N_{\mathcal{C}}(\mathcal{G}(\bar{x}, u)), \tag{4.196}$$

其中

$$\mathcal{L}(x, \mu, u) := f(x, u) + \langle \mu, \mathcal{G}(x, u) \rangle.$$

由 (3.267) 可得, 对 u_0 邻域中的 u 及充分接近于 x_0 的 $\bar{x}(u)$ 与 $\bar{y} := G(\bar{x}(u), u)$,

$$\overline{\lambda}(u) = [D\Xi(\bar{y})]^* \bar{\mu}(u). \tag{4.197}$$

因为 $D_x\mathcal{G}(x_0, u_0)$ 是映上的, 所以 (见命题 4.47 后的注记) 对于 u_0 邻域中的 u_1 与 u_2 及充分接近于 x_0 的 $\bar{x}(u_1)$ 与 $\bar{x}(u_2)$,

$$\|\bar{\mu}(u_1) - \bar{\mu}(u_2)\| = O(\|\bar{x}(u_1) - \bar{x}(u_2)\| + \|u_1 - u_2\|).$$

由 (4.197), 上述性质可传递到问题 (P_u) 的 Lagrange 乘子上去. 所以得到下述结果.

命题 4.82 令 $x_0 \in \mathcal{S}(u_0)$ 是 $G(\cdot, u_0)$ 相对于 K 的非退化点. 则分别存在 u_0 与 x_0 的邻域 N_{u_0} 与 N_{x_0} 及常数 κ 满足若 $u_1, u_2 \in N_{u_0}$, $\bar{x}(u_1) \in \mathcal{S}(u_1) \cap N_{x_0}$, $\bar{x}(u_2) \in \mathcal{S}(u_2) \cap N_{x_0}$, 则

$$\|\overline{\lambda}(u_1) - \overline{\lambda}(u_2)\| \leqslant \kappa(\|\bar{x}(u_1) - \bar{x}(u_2)\| + \|u_1 - u_2\|). \tag{4.198}$$

4.7 Lipschitz 稳定情形的二阶分析

这一节开始最优值函数 $v(u)$ 的二阶方向性质的研究. 可以看出, 在某些情况下, 最优 (接近最优) 解的一阶展开可由最优值函数相应的二阶展开导出.

根据 "接近最优" 解的 Lipschitz 或 Hölder 稳定性与 Lagrange 乘子的存在性, 我们区分三种基本情形. 第一种情形是最优解集是 Lipschitz 稳定的, 则当假设最优解 x_0 唯一时 (见定理 4.25), $v'(u_0, d) = \mathrm{val}(\mathrm{PL}_d)$, 即一阶上估计 (4.51) 是紧的. 进一步, 此种情形可导出最优值函数的二阶上估计, 在某些情形下此上估计是紧的, 能够计算 $o(t^2)$ 最优解的一阶展式.

第二种情形是尽管 Lagrange 乘子集非空, 上估计 (4.15) 不是紧的, 如例 4.23 就是这种情形. 如定理 4.53 所示, 在某些二阶充分条件的弱形式下, 此种情形下的最优解仍然是 $\frac{1}{2}$ 度 Hölder 稳定的. 可行路径类似于 $O(t^{1/2})$ 阶的展开使我们可得到最优值函数更精确的上界, 某些情况下这些上界恰好是紧的. 这些情形下, 可计算 $o(t)$ 最优解的 $O(t^{1/2})$ 阶展式.

第三种情形除了不存在 Lagrange 乘子外, 与第二种情形类似. 此种情形 (非扰动) 问题不是平静的, 因而最优值函数较相应的参数变化要快, 我们证明最优值函数

的变化为参数变化的平方根阶的数量级. 当然可行路径类似于 $O(t^{1/2})$ 的变化的展开式使得我们可估计 $t^{-1/2}(v(u(t))-v(u_0))$ 的上极限. 在某些情况下, 这一上估计是紧的, 由此可计算最优解的 $O(t^{1/2})$ 阶的展式.

这一节研究第一种 (正常) 情形, 此时二阶充分性条件足够强, 可确保最优 (接近最优) 解的 Lipschitz 稳定性, 而第二与第三情形将在 4.8 节中探讨.

除非特别说明, 这一节假设问题 (P_u) 定义为 (4.1) 形式, 其中可行集 $\Phi(u)$ 由形式 (4.2) 的抽象约束给出, X, Y, 与 U 是 Banach 空间, $f(x,u)$ 与 $G(x,u)$ 是二次连续可微的函数. 还假设 $x_0 \in \mathcal{S}(u_0)$ 是给定的最优解, 在 x_0 处沿给定方向 $d \in U$ 的方向正则性条件成立. 考虑下述形式的扰动

$$u(t) := u_0 + td + \frac{1}{2}t^2 r + o(t^2), \quad t \geqslant 0, \tag{4.199}$$

其中 $d, r \in U$ 是 (固定) 方向. 通过导出沿路径 $u(t)$ 的最优值函数的上方估计来开始我们的分析. 然后分别证明两种情况下, 即在方向的广义多面性条件及二阶正则性条件之下, 这一上方估计是紧的.

4.7.1　最优值函数的上方二阶近似

这一节推导最优值函数的上方二阶近似. 考虑由 (4.45) 定义的线性化问题 (PL_d). 令 h 是 (PL_d) 的最优解, 考虑内二阶切集

$$\mathcal{T}_k^2(h,d) := T_k^{i,2}(G(x_0,u_0), DG(x_0,u_0)(h,d)). \tag{4.200}$$

上方估计基于如下形式的抛物 (带余项为 $\varepsilon(t) = o(t^2)$) 路径

$$x(t) := x_0 + th + \frac{1}{2}t^2 w + o(t^2). \tag{4.201}$$

称这一路径是可行的, 若可以选取余项 $\varepsilon(t) = o(t^2)$ 使得对所有的 $t > 0$ 充分小有 $G(x(t), u(t)) \in K$. 注意到, 因为 h 满足 (PL_d) 的约束, 有

$$DG(x_0,u_0)(h,d) \in T_K(G(x_0,u_0)). \tag{4.202}$$

由方向正则性条件及二阶 Taylor 展式, 有

$$\begin{aligned} G(x(t),u(t)) = & G(x_0,u_0) + tDG(x_0,u_0)(h,d) \\ & + \frac{1}{2}t^2[DG(x_0,u_0)(w,r) + D^2G(x_0,u_0)((h,d),(h,d))] + o(t^2), \end{aligned}$$

根据引理 4.10 的上界估计 (4.17) 得, 对于 $o(t^2)$ 项的合适选择, 这样的路径 $x(t)$ 可行的充分必要条件是

$$DG(x_0,u_0)(w,r) + D^2G(x_0,u_0)((h,d),(h,d)) \in \mathcal{T}_K^2(h,d). \tag{4.203}$$

展式 (4.203) 提示我们, 为导出最优值函数方向变化的紧的上方估计, 应该以这样一种方式选取 w: 在满足约束 (4.203) 条件的同时极小化目标函数 Taylor 展式中的二阶项, 即将 w 选取为下述问题的最优解

$$(\mathrm{PQ}_h) \qquad \begin{array}{ll} \min_w & Df(x_0,u_0)(w,r)+D^2f(x_0,u_0)((h,d),(h,d)), \\ \text{s. t.} & DG(x_0,u_0)(w,r)+D^2G(x_0,u_0)((h,d),(h,d))\in \mathcal{T}_K^2(h,d). \end{array}$$

则 h 的最好选取为下述问题的最优解

$$(\mathcal{PQ}) \qquad \min_{h\in\mathcal{S}(\mathrm{PL}_d)} \mathrm{val}(\mathrm{PQ}_h). \tag{4.204}$$

在方向正则性条件之下, 最优值 $\mathrm{val}(\mathrm{PL}_d)$ 有限的充要条件是 Lagrange 乘子集合 $\Lambda(x_0,u_0)$ 非空 (见命题 4.21). 我们得到下述结果.

命题 4.83 *设 $x_0\in\mathcal{S}(u_0)$, Lagrange 乘子集合 $\Lambda(x_0,u_0)$ 非空, 在 x_0 处沿方向 d 的方向正则性条件成立. 则*

$$\limsup_{t\downarrow 0}\frac{v(u(t))-v(u_0)-t\,\mathrm{val}(\mathrm{PL}_d)}{\frac{1}{2}t^2}\leqslant \mathrm{val}(\mathcal{PQ}). \tag{4.205}$$

证明 若 $\mathrm{val}(\mathcal{PQ})=+\infty$, 则 (4.205) 显然成立. 否则, 令 $h\in\mathcal{S}(\mathrm{PL}_d)$ 满足 $\mathrm{val}(\mathrm{PQ}_h)<+\infty$, 即 (PQ_h) 的可行集非空. 令 w 是 (PQ_h) 的可行点. 则 w 满足约束 (4.203), 由上面的讨论知, 存在形式为 (4.201) 的可行路径 $x(t)$, 因此对 $t>0$ 充分小, 有 $v(u(t))\leqslant f(x(t),u(t))$. 用二阶 Taylor 展式

$$\begin{aligned} f(x(t),u(t))=&f(x_0,u_0)+t\,Df(x_0,u_0)(h,d)\\ &+\frac{1}{2}t^2[Df(x_0,u_0)(w,r)+D^2f(x_0,u_0)((h,d),(h,d))]+o(t^2), \end{aligned}$$

由于 $v(u_0)=f(x_0,u_0)$, $\mathrm{val}(\mathrm{PL}_d)=Df(x_0,u_0)(h,d)$, 得到 (4.205) 的左端小于或等于 $\mathrm{val}(\mathrm{PQ}_h)$. 对 $h\in\mathcal{S}(\mathrm{PL}_d)$ 取极小, 不等式 (4.205) 成立. □

如前面指出的那样, 即使 (PL_d) 有有限值, 最优解集 $\mathcal{S}(\mathrm{PL}_d)$ 也可能是空集. 还可能发生对所有 $h\in\mathcal{S}(\mathrm{PL}_d)$, $\mathcal{T}_K^2(h,d)$ 是空集. 这些情况均有 $\mathrm{val}(\mathcal{PQ})=+\infty$, 不等式 (4.205) 简化为平凡的估计 $v(u(t))\leqslant+\infty$. 然而, 另一方面, 若 $\mathrm{val}(\mathcal{PQ})=-\infty$, 则上估计 (4.205) 变为

$$\lim_{t\downarrow 0}\frac{v(u(t))-v(u_0)-t\,\mathrm{val}(\mathrm{PL}_d)}{t^2}=-\infty. \tag{4.206}$$

这是可能发生的, 如当 $v'_+(u_0,d)<\mathrm{val}(\mathrm{PL}_d)$ 时, 见例 4.23. 最后, 若 $\mathrm{val}(\mathcal{PQ})$ 是有限的, 则 (4.205) 变为

$$v(u(t))\leqslant v(u_0)+t\,\mathrm{val}(\mathrm{PL}_d)+\frac{1}{2}t^2\mathrm{val}(\mathcal{PQ})+o(t^2). \tag{4.207}$$

上估计 (4.207) 比由引理 4.57 给出的上界 (4.144) 要精确得多.

问题 (PQ_h) 关于 w 是线性的, 其对偶可类似于 4.3.1 节中 (PL_d) 的对偶那样推导出来. 设内二阶切集 $\mathcal{T}_K^2(h,d)$ 是非空的, $h \in \mathcal{S}(\mathrm{PL}_d)$. 则 (PQ_h) 的对偶可表述为下述形式

$$(\mathrm{DQ}_h) \qquad \max_{\lambda \in \mathcal{S}(\mathrm{DL}_d)} \xi_h(\lambda), \tag{4.208}$$

其中

$$\xi_h(\lambda) := D_u\, L(x_0,\lambda,u_0)r + D^2L(x_0,\lambda,u_0)((h,d)(h,d)) - \sigma(\lambda,\mathcal{T}_K^2(h,d)). \tag{4.209}$$

事实上, 由命题 3.34, $\mathcal{T}_K^2(h,d)$ 的回收锥与下述锥是重合的

$$T_{T_K(y_0)}(DG(x_0,u_0)(h,d)) = \mathrm{cl}\{T_K(y_0) + [\![DG(x_0,u_0)(h,d)]\!]\}, \tag{4.210}$$

因此, 若 $\lambda \notin N_K(y_0)$ 有 $\sigma(\lambda,\mathcal{T}_k^2(h,d))) = +\infty$(这里 $y_0 := G(x_0,u_0)$, $[\![y]\!]$ 记由向量 y 生成的线性空间). 进一步, 若 h 是 (PL_d) 的可行点, 即满足 (4.202), $\lambda \in \Lambda(x_0,u_0)$, 因而 λ 是 (DL_d) 的可行点, 则 (由命题 2.191) 互补条件

$$\langle \lambda, DG(x_0,u_0)(h,d)\rangle = 0 \tag{4.211}$$

成立当且仅当 $h \in \mathcal{S}(\mathrm{PL}_d)$ 与 $\lambda \in \mathcal{S}(\mathrm{DL}_d)$(如 4.47). 所以, (PQ_h) 的对偶的有效域是集 $\mathcal{S}(\mathrm{DL}_d)$, 因此, 由 (PQ_h) 与 (DQ_h) 的对偶性关系可得到.

命题 4.84 设 $x_0 \in \mathcal{S}(u_0)$ 满足集合 $\mathcal{T}_K^2(h,d)$ 与 $\Lambda(x_0,u_0)$ 是非空的, $h \in \mathcal{S}(\mathrm{PL}_d)$, 方向正则性条件成立. 则问题 (PQ_h) 与 (DQ_h) 间不存在对偶间隙, 它们的公共最优值是有限的, (DQ_h) 的最优解集是非空的弱 * 紧致的, 且

$$\sigma(\lambda,\mathcal{T}_K^2(h,d)) \leqslant 0, \quad \forall h \in \mathcal{S}(\mathrm{PL}_d), \quad \forall \lambda \in \mathcal{S}(\mathrm{DL}_d). \tag{4.212}$$

证明 由命题 4.21 得, 在方向正则性条件下, 集合 $\mathcal{S}(\mathrm{PL}_d)$ 是 $\Lambda(x_0,u_0)$ 的非空的弱 * 紧致子集. 进一步, 问题 (PQ_h) 的正则性条件 (2.312) 可以写为下述形式

$$0 \in \mathrm{int}\{\bar{y} + D_xG(x_0,u_0)X - \mathcal{T}_K^2(h,d)\}, \tag{4.213}$$

其中

$$\bar{y} := D_uG(x_0,u_0)r + D^2G(x_0,u_0)((h,d),(h,d)).$$

因为 $\mathcal{T}_K^2(h,d)$ 是非空的, 其回收锥与 (4.210) 式右端给出的锥重合. 于是 (4.213) 右端的集合包含集合

$$y' + D_xG(x_0,u_0)X + [\![D_uG(x_0,u_0)d]\!] - T_K(G(x_0,u_0)), \tag{4.214}$$

其中 $y' \in Y$ 是某一元素. 方向正则性可推出

$$D_xG(x_0,u_0)X + [\![D_uG(x_0,u_0)d]\!] - T_K(G(x_0,u_0)) = Y, \tag{4.215}$$

因此由 (4.214) 给出的集合与 Y 重合. 结果 (4.213) 成立. 由定理 2.165 得问题 (PQ_h) 与 (DQ_h) 间不存在对偶间隙. 因为 $\mathcal{S}(\mathrm{DL}_d)$ 是非空的弱 * 紧致的, 所以 $\mathrm{val}(\mathrm{DQ}_h)$ 是有限的, 进一步, 其最优解集是非空的弱 * 紧致的.

由命题 3.34, 二阶切集 $\mathcal{T}_K^2(h,d)$ 包含在 (4.210) 的右端. 因为任何 $\lambda \in \mathcal{S}(\mathrm{DL}_d)$ 属于 $N_K(y_0)$ 且满足 (4.211), 则可得到 (4.212). □

于是, 若 $\Lambda(x_0,u_0)$ 非空, 存在 $h \in \mathcal{S}(\mathrm{PL}_d)$ 满足 $\mathcal{T}_K^2(h,d)$ 是非空的, 方向正则性成立, 则 $\mathrm{val}(\mathcal{PQ})$ 小于 $+\infty$, 由下述问题的最优值给出

$$(\mathcal{DQ}) \qquad \min_{h\in\mathcal{S}(\mathrm{PL}_d)} \max_{\lambda\in\mathcal{S}(\mathrm{DL}_d)} \xi_h(\lambda), \tag{4.216}$$

其中 $\xi_h(\lambda)$ 由 (4.209) 定义. 由命题 4.83 可得到下述结果.

定理 4.85 令 $x_0 \in \mathcal{S}(u_0)$. 设方向正则性条件成立, $\Lambda(x_0,u_0)$ 非空, 设存在 $h \in \mathcal{S}(\mathrm{PL}_d)$ 满足 $\mathcal{T}_K^2(h,d)$ 非空. 则问题 (4.216) 的最优值 $\mathrm{val}(\mathcal{DQ})$ 小于 $+\infty$, 且

$$\limsup_{t\downarrow 0} \frac{v(u(t)) - v(u_0) - t\ \mathrm{val}(\mathrm{PL}_d)}{\frac{1}{2}t^2} \leqslant \mathrm{val}(\mathcal{DQ}). \tag{4.217}$$

考虑集合

$$\mathcal{T}_K^2(h) := \mathcal{T}_K^2(h,0) = T_K^{i,2}(G(x_0,u_0), D_xG(x_0,u_0)h) \tag{4.218}$$

及下述二阶条件: 存在 $\beta > 0$, 满足对所有的 $h \in C(x_0)\setminus\{0\}$, 不等式

$$\sup_{\lambda\in\mathcal{S}(\mathrm{DL}_d)} \{D_{xx}^2 L(x_0,\lambda,u_0)(h,h) - \sigma(\lambda, \mathcal{T}_K^2(h))\} > 0 \tag{4.219}$$

成立. 这些条件可以通过在 (4.139) 的二阶条件中取 $\mathcal{A}(h) := \mathcal{T}_K^2(h)$ 得到 ((4.139) 用于确保最优解的方向 Lipschitz 稳定性). 在定理 4.85 的假设下, $\mathcal{S}(\mathrm{PL}_d)$ 的回收锥与临界锥 $C(x_0)$ 重合 (见命题 4.21 之后的讨论).

引理 4.86 设 $x_0 \in \mathcal{S}(u_0)$ 满足 Lagrange 乘子集 $\Lambda(x_0,u_0)$ 非空, 方向正则性条件 (4.8) 成立. 且设

(i) 空间 X 是自反的.

(ii) 二阶条件 (4.219) 成立.

(iii) 对所有的 $\lambda \in \mathcal{S}(\mathrm{DL}_d)$, $D_{xx}^2 L(x_0,\lambda,u_0)(\cdot,\cdot)$ 是弱下半连续的, 存在 $\overline{\lambda} \in \mathcal{S}(\mathrm{DL}_d)$, 它是 Legendre 形式.

(iv) 对所有的 $\lambda \in \mathcal{S}(\mathrm{DL}_d)$, 函数 $-\sigma(\lambda, T_K^{i,2}(y_0, \cdot))$ 是弱下半连续的. 则 $\mathrm{val}(\mathcal{DQ}) > -\infty$.

证明　若结论不真, 则存在序列 $h_n \in \mathcal{S}(\mathrm{PL}_d)$ 满足 $\mathrm{val}(\mathrm{PQ}_{h_n}) \to -\infty$. 由命题 4.84, 对所有的 $\lambda \in \mathcal{S}(\mathrm{DL}_d)$, $\sigma(\lambda, \mathcal{T}_K^2(h_n, d)) \leqslant 0$, 且 $\mathrm{val}(\mathrm{PQ}_{h_n}) \geqslant \mathrm{val}(DQ_{h_n})$, 则对非空集合 $\mathcal{S}(\mathrm{DL}_d)$ 中的任何 λ, 有

$$\mathrm{val}(\mathrm{PQ}_{h_n}) \geqslant D_u L(x_0, \lambda, u_0) r + D_{xx}^2 L(x_0, \lambda, u_0)((h_n, d), (h_n, d)).$$

结果, 因为 $\mathcal{S}(\mathrm{DL}_d)$ 是非空的,

$$\mathrm{val}(\mathrm{PQ}_{h_n}) \geqslant -a - b\|h_n\|^2,$$

其中 a 与 b 不依赖于 n. 于是有 $\|h_n\| \to +\infty$. 因为 X 是自反的 Banach 空间, 如有必要可取一子列, 设 $\hat{h}_n := \|h_n\|^{-1} h_n$ 弱收敛到某一 $\bar{h}$, 后者属于 $\mathcal{S}(\mathrm{PL}_d)$ 的回收锥, 即属于临界锥 $C(x_0)$. 根据正齐次性, 由于 $\mathrm{val}(\mathrm{PQ}_{h_n}) \to -\infty$ 与 $\|h_n\| \to +\infty$, 则对每一 $\lambda \in \mathcal{S}(\mathrm{DL}_d)$, 对所有充分大的 n, 下述量值是负的

$$\begin{aligned} &\|h_n\|^{-2}(D^2 L(x_0, \lambda, u_0)((h_n, d), (h_n, d)) - \sigma(\lambda, \mathcal{T}_K^2(h_n, d))) \\ &= D^2 L(x_0, \lambda, u_0)((\hat{h}_n, \|h_n\|^{-1} d), (\hat{h}_n, \|h_n\|^{-1} d)) - \sigma(\lambda, \mathcal{T}_K^2(\hat{h}_n, \|h_n\|^{-1} d)). \end{aligned}$$

由 $D_{xx}^2 L(x_0, \lambda, u_0)(\cdot, \cdot)$ 与 $-\sigma(\lambda, T_K^{i,2}(y_0, \cdot))$ 的弱下半连续性, 因 $\|h_n\| \to +\infty$, 则对所有的 $\lambda \in \mathcal{S}(\mathrm{DL}_d)$,

$$D_{xx}^2 L(x_0, \lambda, u_0)(\bar{h}, \bar{h}) - \sigma(\lambda, \mathcal{T}_K^2(\bar{h})) \leqslant 0.$$

由二阶条件 (4.219) 得 $\bar{h} = 0$, 从而上述等式的左端即 0, 结果

$$\begin{aligned} 0 &= D_{xx}^2 L(x_0, \bar{\lambda}, u_0)(\bar{h}, \bar{h}) \\ &\leqslant \liminf_{n \to \infty} D_{xx}^2 L(x_0, \bar{\lambda}, u_0)((\hat{h}_n, \|h_n\|^{-1} d), (\hat{h}_n, \|h_n\|^{-1} d)) \leqslant 0. \end{aligned}$$

因为 $D_{xx}^2 L(x_0, \bar{\lambda}, u_0)(\cdot, \cdot)$ 是 Legendre 形式, 从而 $\bar{h}$ 是 $\hat{h}_n$ 的强极限, 因为 $\|\hat{h}_n\| = 1$ 但 $\bar{h} = 0$, 这得到矛盾. □

由命题 3.48, 函数 $-\sigma(\lambda, T_K^{i,2}(y_0, \cdot))$ 是凸的. 因此, 它在弱拓扑下是 (弱) 下半连续的当且仅当它在强拓扑下是下半连续的. 由二阶导数的定义, 二次形式 $D_{xx}^2 L(x_0, \lambda, u_0)(\cdot, \cdot)$ 是强连续的. 进一步, 若它是非负值的, 则是凸的, 因此是弱下半连续的.

现在考虑二阶条件

$$\sup_{\lambda \in \mathcal{S}(\mathrm{DL}_d)} \{D_{xx}^2 L(x_0, \lambda, u_0)(h, h) - \sigma(\lambda, \mathcal{T}_K^2(h))\} \geqslant 0, \quad \forall h \in C(x_0), \tag{4.220}$$

它通过把 (4.219) 中的严格正号松弛到非负号得到.

引理 4.87 设 $x_0 \in \mathcal{S}(u_0)$ 满足 Lagrange 乘子集 $\Lambda(x_0, u_0)$ 非空, 设 $\mathrm{val}(\mathcal{PQ}) < +\infty$, 即存在 $h_0 \in \mathcal{S}(\mathrm{PL}_d)$ 满足 $\mathcal{T}_K^2(h_0) \neq \varnothing$. 若二阶条件 (4.220) 不成立, 则 $\mathrm{val}(\mathcal{PQ}) = \mathrm{val}(\mathcal{DQ}) = -\infty$, 因此 (4.206) 成立.

证明 令 h_0 如上, $\bar{h} \in C(x_0)$ 满足

$$\kappa := \sup_{\lambda \in \mathcal{S}(\mathrm{DL}_d)} \{D_{xx}^2 L(x_0, \lambda, u_0)(\bar{h}, \bar{h}) - \sigma(\lambda, \mathcal{T}_K^2(\bar{h}))\}$$

是负值的. 令 $A := DG(x_0, u_0), A_x := D_x G(x_0, u_0)$. 由命题 3.48, 对于任何 $\lambda \in Y^*$, 函数 $\psi_\lambda(\cdot) := -\sigma(\lambda, T_K^{i,2}(y_0, \cdot))$ 是凸的. 由于 $\psi_\lambda(\cdot)$ 是 2 度正齐次的凸的, 对所有的 $t > 0$,

$$\begin{aligned}\psi_\lambda(A(h_0 + t\bar{h}, d)) &= (1+t)^2 \psi_\lambda\left(\frac{1}{1+t} A(h_0, d) + \frac{t}{1+t} A_x \bar{h}\right) \\ &\leqslant (1+t)\psi_\lambda(A(h_0, d)) + t(1+t)\psi_\lambda(A_x \bar{h}).\end{aligned}$$

另一方面, 置 $h_t := h_0 + t\bar{h}$, 得

$$\begin{aligned}D^2 L(x_0, \lambda, u_0)((h_t, d), (h_t, d)) \leqslant\ & (1+t)^2 D^2 L(x_0, \lambda, u_0)((h_0, d), (h_0, d)) \\ & + t(1+t) D_{xx}^2 L(x_0, \lambda, u_0)(\bar{h}, \bar{h}) + O(t),\end{aligned}$$

其中项 $O(t)$ 关于 $\lambda \in \mathcal{S}(\mathrm{DL}_d)$ 是一致的, 因为 $\mathcal{S}(\mathrm{DL}_d)$ 是有界的.

结合 (上述) 两个不等式得到

$$\begin{aligned}&\sup_{\lambda \in \mathcal{S}(\mathrm{DL}_d)} \{D^2 L(x_0, \lambda, u_0)((h_t, d), (h_t, d)) + \psi_\lambda(A(h_t, d))\} \\ &\leqslant (1+t) \sup_{\lambda \in \mathcal{S}(\mathrm{DL}_d)} \{D^2 L(x_0, \lambda, u_0)((h_0, d), (h_0, d)) + \psi_\lambda(A(h_0, d))\} \\ &\quad + t(1+t) \sup_{\lambda \in \mathcal{S}(\mathrm{DL}_d)} \{D_{xx}^2 L(x_0, \lambda, u_0)(\bar{h}, \bar{h}) + \psi_\lambda(A_x \bar{h})\} + O(t) \\ &\leqslant (1+t)\mathrm{val}(\mathrm{PQ}_{h_0}) + t(1+t)\kappa + O(t).\end{aligned}$$

这样可得, 当 $t \to \infty$ 时 $\mathrm{val}(\mathrm{PQ}_{h_t}) \to -\infty$. 因为 $h_t \in \mathcal{S}(\mathrm{PL}_d)$, 得到 $\mathrm{val}(\mathcal{PQ}) = -\infty$, 因而结论成立. □.

在下述广义多面性质条件下, (4.216) 中函数 $\xi_h(\lambda)$ 的 "sigma 项" 消失.

定义 4.88 称参数化问题 (P_u) 在点 (x_0, y_0) 处沿方向 d 满足广义多面性条件, 若

$$0 \in \mathcal{T}_K^2(h, d) := T_K^{i,2}(G(x_0, u_0), DG(x_0, u_0)(h, d)) \tag{4.221}$$

对 $\mathcal{S}(\mathrm{PL}_d)$ 中的 (强) 稠密子集中所有的 h 均成立.

在映射 G 不依赖于 u 的情况, 最优解集 $\mathcal{S}(\mathrm{PL}_d)$ 与临界锥 $C(x_0)$ 相同, 此种情形, 若 (3.120) 中的外二阶切集被相应的内二切集替换, 则上述广义多面性变得与 3.2.3 节 (见定义 3.51) 所讨论的一样.

定义松弛 (relaxed) 问题

$$(\mathrm{DQ}_h^R) \qquad \max_{\lambda \in \mathcal{S}(\mathrm{DL}_d)} \xi_h^R(\lambda), \tag{4.222}$$

$$(\mathcal{DQ}^R) \qquad \min_{h \in \mathcal{S}(\mathrm{PL}_d)} \{\varphi(h) := \sup_{\lambda \in \mathcal{S}(\mathrm{DL}_d)} \xi_h^R(\lambda)\}, \tag{4.223}$$

其中

$$\xi_h^R(\lambda) := D_u L(x_0, \lambda, u_0) r + D^2 L(x_0, \lambda, u_0)((h, d), (h, d)) \tag{4.224}$$

由函数 $\xi_h(\lambda)$ 删除 sigma 项得到.

命题 4.89　设 $\Lambda(x_0, u_0)$ 非空, 沿方向 d 的方向正则性条件成立. 则

(i) 对任何 $h \in \mathcal{S}(\mathrm{PL}_d)$,

$$\mathrm{val}(\mathrm{DQ}_h^R) \leqslant \mathrm{val}(\mathrm{DQ}_h), \tag{4.225}$$

因此, $\mathrm{val}(\mathcal{DQ}^R) \leqslant \mathrm{val}(\mathcal{DQ})$.

(ii) 若方向二阶充分条件 (4.138) 成立, 则 $\mathrm{val}(\mathcal{DQ}^R) > -\infty$.

(iii) 若在点 (x_0, u_0) 处沿方向 d 的广义多面性条件成立, 则 $\mathrm{val}(\mathcal{DQ}^R) = \mathrm{val}(\mathcal{DQ})$.

证明　(i) 由命题 4.84, 对所有 $\lambda \in \mathcal{S}(\mathrm{DL}_d)$ 与 $h \in C(x_0)$, $\sigma(\lambda, \mathcal{T}_K^2(h, d)) \leqslant 0$, 有 $\xi_h^R(\lambda) \leqslant \xi_h(\lambda)$, 从而 $\mathrm{val}(\mathcal{DQ}^R) \leqslant \mathrm{val}(\mathcal{DQ})$.

(ii) 设方向二阶充分条件 (4.138) 成立. 固定 $\eta > 0$, 容易验证, 存在 $\gamma > 0$ 满足若 $h \in \mathcal{S}(\mathrm{PL}_d)$, $\|h\| > \gamma$, 则 $h \in C_\eta(x_0)$. 此种情形, 由 (4.138), 存在 $\beta > 0$, 不依赖于 h, $\lambda \in \mathcal{S}(\mathrm{DL}_d)$ 满足

$$D_{xx}^2 L(x_0, \lambda, u_0)(h, h) \geqslant \beta \|h\|^2.$$

因为 $\mathcal{S}(\mathrm{DL}_d)$ 是有界集合, 由于 $h \in \mathcal{S}(\mathrm{PL}_d)$ 且 $\|h\| > \gamma$, 有

$$\xi_h^R(\lambda) \geqslant \beta \|h\|^2 + c_1 \|h\| + c_2 \geqslant c_3,$$

其中 c_1, c_2 与 c_3 是不依赖于 h 的常数. 由 $(\mathcal{DQ}^R)$ 在有界集上是下方有界的, 得 $\mathrm{val}(\mathcal{DQ}^R) > -\infty$.

(iii) 由 (i), 只需证明 $\mathrm{val}(\mathcal{DQ}) \leqslant \mathrm{val}(\mathcal{DQ}^R)$. 方向正则性条件与广义多面性条件结合命题 4.84 可推出 $\mathrm{val}(\mathcal{DQ}^R) < +\infty$, 因此存在问题 $(\mathcal{DQ}^R)$ 的极小化序列 h_n. 另一方面, 函数

$$\varphi(h) := \sup_{\lambda \in \mathcal{S}(\mathrm{DL}_d)} \xi_h^R(\lambda) \tag{4.226}$$

在 X 的强拓扑下是连续的. 事实上, $\varphi(h)$ 是相对于 X 的强拓扑与 Y^* 的弱 * 拓扑的乘积在集合 $\mathcal{S}(\mathrm{DL}_d)$ 上关于 (h,λ) 的连续函数的上确界, 而集合 $\mathcal{S}(\mathrm{DL}_d)$ 是弱 * 紧致的 (由命题 4.2.1, 这由方向正则性条件得到), $\varphi(h)$ 的连续性由命题 4.4 得到.

设 $\{h_n\}$ 是 $(\mathcal{DQ}^R)$ 的极小化序列, 即 $h_n \in \mathcal{S}(\mathrm{PL}_d)$, $\varphi(h_n) \to \mathrm{val}(\mathcal{DQ}^R)$. 由广义多面性条件, 存在由满足 (4.221) 即 $0 \in \mathcal{T}_K^2(h,d)$ 的点 h 构成的 $\mathcal{S}(\mathrm{PL}_d)$ 的子集 $\mathcal{S}'$, 及 $\mathcal{S}'$ 中的序列 $\{h_n'\}$, 满足 h_n' 与 h_n 充分接近, $\varphi(h_n') \leqslant \varphi(h_n) + n^{-1}$. 于是得到 h_n' 是 $(\mathcal{DQ}^R)$ 的另一极小化序列. 还由于 $0 \in \mathcal{T}_K^2(h_n',d)$, 有 $\sigma(\lambda, \mathcal{T}_K^2(h_n',d)) \geqslant 0$, 连同另一不等式 (4.212) 可推出 $\sigma(\lambda, \mathcal{T}_K^2(h_n',d)) = 0$ 对所有 $\lambda \in \mathcal{S}(\mathrm{DL}_d)$ 成立. 于是有

$$\mathrm{val}(\mathcal{DQ}) \leqslant \lim_{n\to\infty} \varphi(h_n') = \mathrm{val}(\mathcal{DQ}^R),$$

证毕. □

在定理 4.85 的条件下, $\mathrm{val}(\mathcal{DQ}) < +\infty$ 且二阶条件 (4.220) 是 $\mathrm{val}(\mathcal{DQ}) > -\infty$ 的必要条件. 作为特殊情况, 若对所有的 $h \in C(x_0)$, (4.220) 中的 sigma 项消失, 则下述条件是 $\mathrm{val}(\mathcal{DQ}) > -\infty$ 的必要条件

$$\sup_{\lambda \in \mathcal{S}(\mathrm{DL}_d)} D_{xx}^2 L(x_0,\lambda,u_0)(h,h) \geqslant 0, \quad \forall h \in C(x_0). \tag{4.227}$$

下述结果类似于命题 3.54.

命题 4.90 *设沿方向 $d \in U$ 的方向正则性条件成立, 映射 $D_xG(x_0,u_0): X \mapsto Y$ 是映上的, K 在 $y_0 := G(x_0,u_0)$ 处是多面性的, $\Lambda(x_0,u_0)$ 非空. 则存在 $\mathcal{S}(\mathrm{PL}_d)$ 的稠密子集 S' 满足对每一 $h \in S'$, $DG(x_0,u_0)(h,d) \in \mathcal{R}_K(y_0)$, 因此 P_u 在 (x_0,u_0) 沿方向 d 满足广义的多面性条件.*

证明 因为 $\Lambda(x_0,u_0)$ 非空, 由方向正则性条件, 问题 (DL_d) 有最优解 λ, 问题 (PL_d) 与 (DL_d) 间不存在对偶间隙 (见命题 4.2.1). 则 $h \in \mathcal{S}(\mathrm{PL}_d)$ 当且仅当 $DG(x_0,u_0)(h,d) \in T_K(y_0)$ 且

$$\langle \lambda, DG(x_0,u_0)(h,d) \rangle = 0$$

(见 (4.47)). 即

$$A(\mathcal{S}(\mathrm{PL}_d)) + b = T_K(y_0) \cap (\ker\lambda), \tag{4.228}$$

其中 $A := D_xG(x_0,u_0), b := D_uG(x_0,u_0)d$. 因 λ 是 Lagramge 乘子, 有 $\lambda \in N_K(y_0)$. 所以由 K 是于 y_0 处为多面性的, 集合 $T_K(y_0) \cap (\ker\lambda)$ 有稠密子集 $D \subset \mathcal{R}_k(y_0)$. 由于 A 是映上的, 由开映射定理, $S' := A^{-1}(D-b)$ 在 $\mathcal{S}(\mathrm{PL}_d)$ 中稠密. 最后, 对任何 $h \in S'$, $DG(x_0,u_0)(h,d) \in \mathcal{R}_K(y_0)$, 因此 (4.221) 成立. □.

注意到, 上述证明中用到的方向正则性条件仅是为保证 (PL_d) 与 (DL_d) 间没有对偶间隙, 且 (DL_d) 有最优解. 在约束映射 G 不依赖于 u 的情况, 这一点 (若 $(\Lambda x_0, u_0)$ 非空) 是自动成立的, 因此上述结果与命题 3.54 中的结果相同.

4.7.2　没有 sigma 项的下方估计

这一节研究最优值函数的下方二阶近似, 这种讨论类似于上一节关于上方近似的做法, 但不具有 "sigma 项". 正如命题 4.89 中所证明的那样, 在广义多面性成立的条件下 sigma 项消失. 此种情况, 上方与下方估计是重合的, 这使得我们可以推导最优解的一阶展式.

下方估计基于下述不等式

$$f(x,u)-f(x_0,u_0)\geqslant L(x,\lambda,u)-L(x_0,\lambda,u_0). \tag{4.229}$$

因为 $\Lambda(x_0,u_0)\subset N_K(G(x_0,u_0))$, 当 $\lambda\in\Lambda(x_0,u_0)$, $G(x,u)\in K$, 即 $x\in\Phi(u)$ 时, 上述不等式成立.

考虑上一节定义的问题 (DQ_h^R), $(\mathcal{D}Q^R)$, 函数 $\xi_h^R(\lambda)$ 与 $\varphi(h)$(见 (4.222) ~ (4.224)). 现在推导最优值函数的下方估计. 注意到, 在下面的定理中, 二阶充分条件并不显式地被提及. 然而, 它们通过设 $o(t^2)$ 最优解 $\bar{x}(t)$ 的 Lipschitz 稳定性而被隐式地用到, 这往往通过验证二阶条件的某种强形式 (见定理 4.55) 体现. 可参看下述定理证明之后的注记.

定理 4.91　*设 X 是自反的 Banach 空间. 设*

(i) $\Lambda(x_0,u_0)$ 是非空的且方向正则性条件成立.

(ii) 问题 $(\mathrm{P}_{u(t)})$ 具有 $o(t^2)$ 最优解 $\bar{x}(t)$ 满足 $\|\bar{x}(t)-x_0\|=O(t)$, $t\geqslant 0$.

(iii) 对所有的 $\lambda\in\Lambda(x_0,u_0)$, 函数 $D_{xx}^2L(x_0,\lambda,u_0)(\cdot,\cdot)$ 是弱下半连续的.

则

$$\liminf_{t\downarrow 0}\frac{v(u(t))-v(u_0)-t\,\mathrm{val}(\mathrm{PL}_d)}{\frac{1}{2}t^2}\geqslant \mathrm{val}(\mathcal{D}Q^R). \tag{4.230}$$

证明　考虑比率 $h(t):=t^{-1}(\bar{x}(t)-x_0)$. 因为假设 $\bar{x}(t)$ 是 Lipschitz 稳定的, 对所有的 $t>0$ 充分小, 有 $h(t)$ 是有界的. 当 $t\downarrow 0$ 时, 令 $\bar{h}$ 是 $h(t)$ 的弱极限点, 序列 $t_n\downarrow 0$ 满足 (4.230) 左端的 lim inf 沿此序列被取得. 注意到, 因为 X 是自反的, $h(t)$ 是有界的, 这样的弱极限点是存在的. 因为 $\bar{x}(t)$ 是 $(\mathrm{P}_{u(t)})$ 的可行点, 即 $G(\bar{x}(t),u(t))\in K$, 有

$$T_K(G(x_0,u_0))\ni\frac{G(\bar{x}(t),u(t))-G(x_0,u_0)}{t}=DG(x_0,u_0)(h(t),d)+o(1).$$

因为 $T_K(G(x_0,u_0))$ 是闭的凸的, 所以是弱闭的, 于是得到 $\bar{h}$ 属于问题 (PL_d) 的可行集. 又由于 $\bar{x}(t)$ 是 $o(t^2)$ 最优的, 有

$$\begin{aligned} v(u(t)) &= f(\bar{x}(t),u(t)) + o(t^2) \\ &= v(u_0) + t\ Df(x_0,u_0)(h(t),d) + o(t), \end{aligned}$$

由定理 4.25, $v(u(t)) = v(u_0) + t\ \mathrm{val}(\mathrm{PL}_d) + o(t)$, 得到 $\bar{h} \in \mathcal{S}(\mathrm{PL}_d)$.

令 $\lambda \in \mathcal{S}(\mathrm{DL}_d)$. 则由 (4.229),

$$\begin{aligned} v(u(t)) &= f(\bar{x}(t),u(t)) + o(t^2) \\ &\geqslant f(x_0,u_0) + L(\bar{x}(t),\lambda,u(t)) - L(x_0,\lambda,u_0) + o(t^2) \\ &= v(u_0) + tD_uL(x_0,\lambda,u_0)d + \frac{1}{2}t^2D_uL(x_0,\lambda,u_0)r \\ &\quad + \frac{1}{2}t^2D^2L(x_0,\lambda,u_0)((h(t),d),(h(t),d)) + o(t^2) \\ &= v(u_0) + t\ \mathrm{val}(\mathrm{DL}_d) + \frac{1}{2}t^2\xi^R_{h(t)}(\lambda) + o(t^2). \end{aligned} \tag{4.231}$$

根据假设 (iii), $\xi^R_h(\lambda)$ 是 h 的弱下半连续函数. 进一步, 由于 $\mathrm{val}(\mathrm{DL}_d) = \mathrm{val}(\mathrm{PL}_d)$,

$$\liminf_{t\downarrow 0} \frac{v(u(t)) - v(u_0) - t\ \mathrm{val}(\mathrm{PL}_d)}{\frac{1}{2}t^2} \geqslant \xi^R_{\bar{h}}(\lambda). \tag{4.232}$$

对上述不等式的右端关于 $\lambda \in \mathcal{S}(\mathrm{DL}_d)$ 取极大化, 得

$$\liminf_{t\downarrow 0} \frac{v(u(t)) - v(u_0) - t\ \mathrm{val}(\mathrm{PL}_d)}{\frac{1}{2}t^2} \geqslant \varphi(\bar{h}), \tag{4.233}$$

其中 $\varphi(\cdot)$ 由 (4.226) 定义. 因为 $\bar{h} \in \mathcal{S}(\mathrm{PL}_d)$, 得到 (4.230). □

注 4.92 由 (4.231) 可得, 若存在 $(\mathrm{P}_{u(t)})$ 的在 x_0 处 Lipschitz 稳定的 $o(t^2)$ 最优解 $\bar{x}(t)$, 则

$$v(u(t)) \geqslant v(u_0) + t\ \mathrm{val}(\mathrm{DL}_d) + O(t^2). \tag{4.234}$$

这是因为 $h(t) := t^{-1}(\bar{x}(t) - x_0)$ 有界, $D^2L(x_0,\lambda,u_0)((\cdot,d),(\cdot,d))$ 是连续的二次函数, 它在有界集上是有界的. 在方向正则性条件成立时, $\mathrm{val}(\mathrm{DL}_d) = \mathrm{val}(\mathrm{PL}_d)$. 所以, 上述不等式 (4.234) 与定理 4.85 中的不等式 (4.217) 只有在 $\mathrm{val}(\mathcal{DQ}) > -\infty$ 的前提下才可能同时成立. 由引理 4.87, 若 $\mathrm{val}(\mathcal{DQ}) > -\infty$, 则二阶条件 (4.220) 成立. 因此, 二阶条件 (4.220) 是 $\bar{x}(t)$ 的 Lipschitz 稳定性的必要性条件. 在这一方面, 可将

条件 (4.220) 与二阶充分条件 (4.139) 相比较, 后者用于建立 (接近) 最优解的方向 Lipschitz 稳定性.

注 4.93　Lipschitz 稳定的 $(\mathrm{P}_{u(t)})$ 的 $o(t^2)$ 最优解路径 $\bar{x}(t)$ 的存在性假设依赖于路径 $u(t)$ 的特定选择. 然而, 在定理 4.55 的假设下, 只要存在收敛到点 x_0 的问题 $(\mathrm{P}_{u(t)})$ 的 $o(t^2)$ 最优解路径, 这样一个路径就存在.

通常而言, 定理 4.91 中的下估计与定理 4.85 的上估计间存在着间隙, 这是由于下估计 (4.230) 没有考虑到集合 K 的可能的曲率, 即它不包含 "sigma 项"$\sigma(\lambda, \mathcal{T}_K^2(h,d))$. 在命题 4.89 中已经证明, 在方向的广义多面性条件之下 (见定理 4.88), 问题 $(\mathcal{DQ})$ 中的 "sigma" 项消失.

将定理 4.85、命题 4.89 与定理 4.91 结合起来, 得到下述结果.

定理 4.94　设 X 是自反的 Banach 空间, 且设

(i) $\Lambda(x_0,u_0)$ 非空且方向正则性条件成立.

(ii) $(P_{u(t)})$ 具有 $o(t^2)$ 最优解 $\bar{x}(t)$ 满足 $\|\bar{x}(t)-x_0\|=O(t)$, $t\geqslant 0$.

(iii) 对所有的 $\lambda\in\Lambda(x_0,u_0)$, 函数 $D_{xx}^2L(x_0,\lambda,u_0)(\cdot,\cdot)$ 是弱下半连续的.

(iv) 方向的增广的多面性条件满足.

则

$$\lim_{t\downarrow 0}\frac{v(u(t))-v(u_0)-t\ \mathrm{val}(\mathrm{PL}_d)}{\dfrac{1}{2}t^2}=\mathrm{val}(\mathcal{DQ}^R). \tag{4.235}$$

假设上述定理的结论成立, 现在研究最优与 "接近" 最优解的一阶方向性质.

定理 4.95　设 X 是自反的 Banach 空间, 且设

(i) $\Lambda(x_0,u_0)$ 非空且方向正则性条件成立.

(ii) 等式 (4.235) 成立.

(iii) 对所有的 $\lambda\in\Lambda(x_0,u_0)$, 函数 $D_{xx}^2L(x_0,\lambda,u_0)(\cdot,\cdot)$ 是弱下半连续的.

则下述结论成立:

(a) 若 $\bar{x}(t)$ 是 $(\mathrm{P}_{u(t)})$ 的 Lipschitz 稳定的 $o(t^2)$ 最优解, 则 $t^{-1}(\bar{x}(t)-x_0)$ 的每一弱极限点均属于 $\mathcal{S}(\mathcal{DQ}^R)$.

(b) 若还有, 对每一 $\lambda\in\mathcal{S}(\mathrm{DL}_d)$, $D_{xx}^2L(x_0,\lambda,u_0)(\cdot,\cdot)$ 是 Legendre 形式, 则 $t^{-1}(\bar{x}(t)-x_0)$ 的每一弱极限点也是强极限点, 若 $(\mathcal{DQ}^R)$ 有唯一的最优解 $\bar{h}$, 则

$$\bar{x}(t)=x_0+t\bar{h}+o(t),\quad t\geqslant 0. \tag{4.236}$$

(c) 若 $h\in\mathcal{S}(\mathcal{DQ}^R)$, $w\in\mathcal{S}(PQ_h)$, 则存在 $(\mathrm{P}_{u(t)})$ 的 $o(t^2)$ 最优解路径, 具有形式 $x(t)=x_0+th+\dfrac{1}{2}t^2w+o(t^2)$.

证明 (a) 令 $\bar{h}$ 是 $t^{-1}(\bar{x}(t)-x_0)$ 的弱极限点, 即存在 $t_n \downarrow 0$, 有 $h_n \xrightarrow{w} \bar{h}$, 其中 $h_n := t_n^{-1}(\bar{x}(t_n)-x_0)$. 于是, 如定理 4.91 证明的那样, 对 $\lambda \in \Lambda(x_0,u_0)$ 有

$$v(u(t_n)) \geqslant v(u_0) + t_n \ \mathrm{val}(\mathrm{PL}_d) + \frac{1}{2}t_n^2\xi_{h_n}^R(\lambda) + o(t_n^2), \tag{4.237}$$

且 $\bar{h} \in \mathcal{S}(\mathrm{PL}_d)$. 由假设 (iii), $\xi_h^R(\lambda)$ 关于 h 是弱下半连续的. 从而得到

$$\liminf_{n\to\infty} \frac{v(u(t_n)) - v(u_0) - t_n \ \mathrm{val}(\mathrm{PL}_d)}{\frac{1}{2}t_n^2} \geqslant \xi_{\bar{h}}^R(\lambda).$$

在上面的不等式右端相对于 $\lambda \in \Lambda(x_0,u_0)$ 取最大化, 得到左端大于或等于 $\varphi(\bar{h})$. 由 (ii), 上述不等式的左端等于 $\mathrm{val}(\mathcal{D}Q^R)$ 且 $\bar{h} \in \mathcal{S}(\mathrm{PL}_d)$, 得到 $\bar{h}$ 即问题 $(\mathcal{D}Q^R)$ 的最优解.

(b) 在 (4.237) 中对有界集 $\mathcal{S}(\mathrm{DL}_d)$ 取 λ 的极大化, 得到

$$v(u(t_n)) \geqslant v(u_0) + t_n \ \mathrm{val}(\mathrm{PL}_d) + \frac{1}{2}t_n^2\varphi(h_n) + o(t_n^2).$$

则由 (4.235) 有 $\varphi(\bar{h}) \geqslant \limsup\limits_{n\to\infty}\varphi(h_n)$. 另一方面, ξ_h^R 关于 h 是弱下半连续的. 从而 $\varphi(\cdot)$ 也是弱下半连续的. 结果, $\varphi(\bar{h}) \leqslant \liminf\limits_{n\to\infty}\varphi(h_n)$, 从而有 $\varphi(h_n) \to \varphi(\bar{h})$. 因为 $\mathcal{S}(\mathrm{DL}_d)$ 是弱 * 紧致的集合, 存在 $\overline{\lambda} \in \mathrm{argmax}\left\{\xi_{\bar{h}}^R(\lambda) : \lambda \in \mathcal{S}(\mathrm{DL}_d)\right\}$. 取 $\lambda = \overline{\lambda}$, 由 (4.237) 得

$$\lim_{n\to\infty}\varphi(h_n) \geqslant \limsup_{n\to\infty}\xi_{h_n}^R(\overline{\lambda}) \geqslant \liminf_{n\to\infty}\xi_{h_n}^R(\overline{\lambda}) \geqslant \xi_{h_n}^R(\overline{\lambda}) = \varphi(\bar{h}),$$

因而有 $\xi_{h_n}^R(\overline{\lambda}) \to \xi_{\bar{h}}^R(\overline{\lambda})$. 因为函数 $D_{xx}^2L(x_0,\overline{\lambda},u_0)(\cdot,\cdot)$ 是 Legendre 形式, 有 $h_n \to \bar{h}$ 按强拓扑成立. 若还有 $(\mathcal{D}Q^R)$ 有唯一最优解, 则 $h(t)$ 收敛于这个解, 这意味着 (4.236) 成立.

(c) 路径 $x(t) := x_0 + th + \frac{1}{2}t^2w$ 满足 $\mathrm{dist}(G(x(t),u(t)),K) = o(t^2)$. 由引理 4.10, 存在可行的路径具有形式 $x(t)+\varepsilon(t)$, $\|\varepsilon(t)\| = o(t^2)$. 将 $f(x(t),u(t))$ 展至二阶, 得到

$$f(x(t),u(t)) = f(x_0,u_0) + t \ \mathrm{val}(\mathrm{PL}_d) + \frac{1}{2}t^2\mathrm{val}(\mathrm{PQ}_h) + o(t^2).$$

结果得证. □

注 4.96 不难给出确保 $(\mathcal{D}Q^R)$ 有唯一最优解的 (二阶) 充分性条件. 例如, 若

$$D_{xx}^2L(x_0,\lambda,u_0)(h,h) > 0, \quad \forall h \in \mathrm{Sp}\ (\mathcal{S}(\mathrm{PL}_d)), \quad \forall\lambda \in \mathcal{S}(\mathrm{DL}_d)^{①} \tag{4.238}$$

则 $\varphi(\cdot)$ 是由 $\mathcal{S}(\mathrm{PL}_d)$ 生成的线性空间上的严格凸函数的上确界, 因而它在 $\mathcal{S}(\mathrm{PL}_d)$ 上有唯一的极小点.

① 式 (4.238) 加条件 $h \neq 0$ 才可以.

注 4.97　若空间 X 是有限维的，上述定理中的条件当然是可以简化的. 此种情形，弱拓扑与通常 (强) 拓扑重合，$D^2_{xx}L(x_0,\lambda,u_0)(\cdot,\cdot)$ 的弱下半连续性可由假设 $f(x,u)$ 与 $G(x,u)$ 二次连续可微性推得. 此种情况下，$D^2_{xx}L(x_0,\lambda,u_0)(\cdot,\cdot)$ 是 Legendre 形式的假设也是自然成立的.

注 4.98　若 X 是有限维的，像 (2.174) 那样，可行集合由有限多个约束定义，则 (P_u) 成为非线性规划的标准问题. 此种情形，广义多面性条件是自动成立的，sigma 项消失. 还有集合 $\Lambda(x_0,u_0)$ 是闭的凸多胞体，因而有有限个极点(顶点). 集合 $\mathcal{S}(\mathrm{DL}_d)$ 是此多胞体的一个面. 有趣的是，若 $\Lambda(x_0,u_0)$ 是非单点集合，$D_uG(x_0,u_0)\neq 0$，则 $\mathcal{S}(\mathrm{DL}_d)$ 以非连续的方式依赖于方向 d. 集合 $\mathcal{S}(\mathrm{PL}_d)$ 由下述线性规划的最优解构成

$$\begin{aligned}\min_{h\in\mathbb{R}^n}\ & h\cdot\nabla_x f(x_0,u_0)\\ \text{s. t.}\ & h\cdot\nabla_x g_i(x_0,u_0)+d\cdot\nabla_u g_i(x_0,u_0)=0,\quad i=1,\cdots,q,\\ & h\cdot\nabla_x g_i(x_0,u_0)+d\cdot\nabla_u g_i(x_0,u_0)\leqslant 0,\quad i\in I(x_0,u_0),\end{aligned}$$

其中 $I(x_0,u_0)$ 记在 (x_0,u_0) 处起作用的不等式约束的指标集.

进一步，在方向正则性、$\Lambda(x_0,u_0)$ 非空与二阶条件 (4.238) 这些假设之下，$\mathrm{val}(\mathrm{PL}_d)$ 是有限的，$\mathcal{S}(\mathrm{DL}_d)$ 是非空的 (因为它是线性规划问题的最优解集)，若 $\bar{x}(t)$ 是 $(\mathrm{P}_{u(t)})$ 的 $o(t^2)$ 最优解且收敛于 x_0，则一阶展式 (4.236) 成立，其中 $\bar{h}$ 是 $(\mathcal{D}Q^R)$ 的唯一最优解. 若还有 $\mathcal{S}(\mathrm{DL}_d)$ 是单点集，则 $(\mathcal{D}Q^R)$ 变为一个二次规划问题. 进一步，若 $\Lambda(x_0,u_0)$ 是单点集，且对应于起作用的不等式约束的所有的 Lagrange 乘子均是严格正的 (严格互补松弛条件)，则 $\mathcal{S}(\mathrm{PL}_d)$ 成为仿射空间，$(\mathcal{D}Q^R)$ 的最优解集可以表示为闭形式.

例 4.99　考虑问题

$$\min_{(x_1,x_2)\in\mathbb{R}^2}\frac{1}{2}(x_1-1)^2+\frac{1}{2}x_2^2\quad \text{s.t.}\quad x_1\leqslant 0,\quad x_1+u_1x_2+u_2\leqslant 0, \tag{4.239}$$

它依赖于参数向量 $(u_1,u_2)\in\mathbb{R}^2$. 对 $u_0=(0,0)$，这一问题有唯一的解 $x_0=(0,0)$. Mangasarian-Fromovitz 约束规范在 x_0 处成立，相应的 Lagrange 乘子集合 Λ_0 是

$$\Lambda_0=\{(\lambda_1,\lambda_2):\lambda_1+\lambda_2=1,\lambda_1\geqslant 0,\lambda_2\geqslant 0\}.$$

强二阶充分条件亦是成立的，因而定理 4.95 的结论可以应用. 有[①]

$$D^2L(x_0,\lambda,u_0)((h,d),(h,d))=h_1^2+h_2^2+\lambda_2 d_1 h_2$$

且 $\mathcal{S}(\mathrm{PL}_d)$ 是下述线性化问题的最优解

$$\min\ (-h_1)\quad \text{s.t.}\ \ h_1\leqslant 0,\quad h_1+d_2\leqslant 0. \tag{4.240}$$

① 译者计算为 $h_1^2+h_2^2+2\lambda_2 d_1 h_2$.

进一步, $D_uL(x_0,\lambda,u_0)=(0,\lambda_2)$, 因此 $\mathcal{S}(\mathrm{DL}_d)$ 由 $d_2\lambda_2$ 在 Λ_0 上的极大值点构成.

令 $d=(1,0), r=0$. 则 $\mathcal{S}(\mathrm{PL}_d)=\{0\}\times\mathbb{R}, \mathcal{S}(\mathrm{DL}_d)=\Lambda_0$, 且 $\varphi(h)=\max\{h_1^2+h_2^2, h_1^2+h_2^2+2h_2\}$①. 此种情形有 $\bar{h}=(0,0)$, 从而 $\bar{x}'(0,d)=(0,0)$. 令 $d=(1,\gamma)$, 其中 $\gamma>0$ 是某一正数. 则 $\mathcal{S}(\mathrm{PL}_d)=\{-\gamma\}\times\mathbb{R}, \mathcal{S}(\mathrm{DL}_d)=\{(0,1)\}, \varphi(h)=h_1^2+h_2^2+2h_2$. 此种情况有 $\bar{h}=(-\gamma,-1)$, 因而 $\bar{x}'(0,d)=(-\gamma_1,-1)$. 在这个例子中可以看到, 最优解 $\bar{x}(u)$ 在 $u=0$ 处是方向可微的, 但方向导数 $\bar{x}'(0,d)$ 是方向 d 的非连续函数. 于是, $\bar{x}(u)$ 不是 Hadamard 意义下 (见命题 2.46) 方向可微的, 且在 $u=0$ 的任何邻域上均不是 Lipschitz 连续的 (见命题 2.49). $\bar{x}'(0,\cdot)$ 的这一非连续性质的原因是最优解集 $\mathcal{S}(\mathrm{DL}_d)$ 是以非连续方式依赖于 d 的.

4.7.3 二阶正则情形

填补最优值函数的上二阶近似与下二阶近似之间间隙的另一种可能情形是集合 K 二阶正则性条件成立. 在下述定理中, 用上二阶近似集合的概念 (见定义 3.28) 导出一下方近似. 用 $D^2L(x,\lambda,u)$ 记 Lagrange 函数关于 (x,u) 的二阶导数.

定理 4.100 设

(i) 空间 X 是有限维的.

(ii) $\mathcal{S}(u_0)=\{x_0\}, \Lambda(x_0,u_0)\neq\varnothing$.

(iii) 沿方向 d 的方向正则性条件成立.

(iv) 存在 $(P_{u(t)})$ 的 $o(t^2)$ 最优解 $\bar{x}(t)$, 在 x_0 处为 Lipschitz 稳定的.

(v) 每一 $h\in\mathcal{S}(\mathrm{PL}_d)$ 对应点 $y_0:=G(x_0,u_0)$ 处沿方向 $DG(x_0,u_0)(h,d)$ 关于线性映射 $M:=D_xG(x_0,u_0)$ 的 K 的上二阶近似集 $\mathcal{A}(h,d)$.

则

$$\liminf_{t\downarrow 0}\frac{v(u(t))-v(u_0)-t\,\mathrm{val}(\mathrm{PL}_d)}{\frac{1}{2}t^2}\geqslant\vartheta(d,r), \tag{4.241}$$

其中 $\vartheta(d,r)$ 是下述极小 - 极大问题的最优值

$$\min_{h\in\mathcal{S}(\mathrm{PL}_d)}\ \max_{\lambda\in\mathcal{S}(\mathrm{DL}_d)}\ \theta_h(\lambda), \tag{4.242}$$

$$\theta_h(\lambda):=D_uL(x_0,\lambda,u_0)r+D^2L(x_0,\lambda,u_0)((h,d),(h,d))-\sigma(\lambda,\mathcal{A}(h,d)). \tag{4.243}$$

证明 考虑序列 $t_n\downarrow 0$, 它满足 (4.241) 左端的下极限取到沿这一序列的极限值. 置

$$x_n:=\bar{x}(t_n),\quad h_n:=t_n^{-1}(x_n-x_0),\quad u_n:=u(t_n).$$

① 从此表达式看出, 确实是 $2\lambda_2 d_1 h_2$.

因为 $\bar{x}(t)$ 是 Lipschitz 稳定的, 所以序列 $\{h_n\}$ 是有界的, 从而如有必要, 可抽取它的子列, 不妨设 $\{h_n\}$ 收敛到某一 $h \in X$. 由 $G(x_n, u_n)$ 与 $f(x_n, u_n)$ 各自的一阶展式, 得到 $DG(x_0, u_0)(h, d) \in T_K(G(x_0, u_0))$, 且[①]

$$\begin{aligned} v(u_n) - v(u_0) &= f(x_n, u_n) - f(x_0, u_0) + o(t_n^2) \\ &= t_n Df(x_0, u_0)(h, d) + o(t_n) \end{aligned}$$

结合命题 4.22 中的不等式 (4.51) 表明

$$Df(x_0, u_0)(h, d) \leqslant \limsup_{n\to\infty} \frac{v(u_n) - v(u_0)}{t_n} \leqslant \mathrm{val}(\mathrm{PL}_d).$$

因为 h 是 (PL_d) 的可行点, 所以 $h \in \mathcal{S}(\mathrm{PL}_d)$, 于是有 $x_n = x_0 + t_n h + \frac{1}{2}t_n^2 w_n$, 其中 $t_n w_n \to 0$. 可以得到

$$\begin{aligned} G(x_n, u_n) = y_0 &+ t_n DG(x_0, u_0)(h, d) + \frac{1}{2}t_n^2 DG(x_0, u_0)(w_n, r) \\ &+ \frac{1}{2}t_n^2 D^2G(x_0, u_0)((h, d), (h, d)) + o(t_n^2). \end{aligned}$$

因为 $\mathcal{A}(h, d)$ 是上二阶近似集合, 可得

$$DG(x_0, u_0)(w_n, r) + D^2G(x_0, u_0)((h, d), (h, d)) + o(1) \in \mathcal{A}(h, d).$$

令 $\lambda \in \mathcal{S}(\mathrm{DL}_d)$. 因为 $\langle \lambda, DG(x_0, u_0)(h, d)\rangle = 0$, 置 $y_n := G(x_n, u_n)$, 导出[②]

$$\begin{aligned} \langle \lambda, y_n - y_0\rangle &= t_n^2 \langle \lambda, DG(x_0, u_0)(w_n, r) + D^2G(x_0, u_0)((h, d), (h, d))\rangle, \\ &\leqslant t_n^2[\sigma(\lambda, \mathcal{A}(h, d)) + o(1)] \end{aligned}$$

由于 $\mathcal{S}(\mathrm{DL}_d)$ 是有界的, $o(1)$ 项关于 $\lambda \in \mathcal{S}(\mathrm{DL}_d)$ 一致收敛到零. 结果

$$\begin{aligned} v(u_n) - v(u_0) &= f(x_n, u_n) - f(x_0, u_0) + o(t_n^2) \\ &= L(x_n, \lambda, u_n) - L(x_0, \lambda, u_0) - \langle \lambda, y_n - y_0\rangle + o(t_n^2) \\ &\geqslant t_n D_u L(x_0, \lambda, u_0)d + \frac{1}{2}t_n^2 D_u L(x_0, \lambda, u_0)r \\ &\quad + \frac{1}{2}t_n^2[D^2L(x_0, \lambda, u_0)((h, d), (h, d)) - \sigma(\lambda, \mathcal{A}(h, d))] + o(t_n^2). \end{aligned}$$

因为 $\lambda \in \mathcal{S}(\mathrm{DL}_d)$, 所以 $D_u L(x_0, \lambda, u_0)d = \mathrm{val}(\mathrm{DL}_d) = \mathrm{val}(\mathrm{PL}_d)$. 由于 $\lambda \in \mathcal{S}(\mathrm{DL}_d)$ 是任意的, $h \in \mathcal{S}(\mathrm{PL}_d)$, 可推出 (4.241). □

如果对每一 $h \in \mathcal{S}(\mathrm{PL}_d)$, 因 (4.200) 定义的内二阶切集 $T_K^2(h, d)$ 可作为上二阶近似集合, 则下方估计 (4.241) 与定理 4.85 给出的上方近似 (4.217) 重合, 即在附加

① 第一式中 $o(t_n^2)$ 应为 $o(t_n)$.

② 原著中 σ 前没有 t_n^2.

的 K 的二阶正则性假设下, 估计 (4.241) 与 (4.217) 间的间隙消失, 则得到最优值函数的下述二阶展开式.

定理 4.101 设

(i) 空间 X 是有限维的.

(ii) $\mathcal{S}(u_0) = \{x_0\}$, $\Lambda(x_0, u_0) \neq \varnothing$.

(iii) 沿方向 d 的方向正则性条件成立.

(iv) 存在 $(\mathrm{P}_{u(t)})$ 的 $o(t^2)$ 最优解 $\bar{x}(t)$, 在 x_0 处是 Lipschitz 稳定的.

(v) 对每一 $h \in \mathcal{S}(\mathrm{PL}_d)$, 集合 K 在点 $G(x_0, u_0)$ 处沿方向 $DG(x_0, u_0)(h, d)$ 相对于线性映射 $D_xG(x_0, u_0)$ 是二阶正则的.

则下述结论成立:

(a) 问题 $(\mathcal{DQ})$(由 (4.216) 定义) 的最优值是有限的, 对 $t \geqslant 0$,

$$v(u(t)) = v(u_0) + t\ \mathrm{val}(\mathrm{PL}_d) + \frac{1}{2}t^2\ \mathrm{val}(\mathcal{DQ}) + o(t^2). \tag{4.244}$$

(b) 对 $(\mathrm{P}_{u(t)})$ 的任何 $o(t^2)$ 最优解 $\bar{x}(t)$, 当 $t \downarrow 0$ 时 $(\bar{x}(t) - \bar{x})/t$ 的每一聚点均是问题 $(\mathcal{DQ})$ 的最优解.

(c) 设 $\bar{h}$ 是 $(\mathcal{DQ})$ 的最优解, $\bar{w}$ 是 (PQ_h) 的相应的最优解(假设这样的最优解确实存在). 则存在 $(\mathrm{P}_{u(t)})$ 的具有形式 $\hat{x}(t) := x_0 + t\bar{h} + \frac{1}{2}t^2\bar{w} + o(t^2)$ 的 $o(t^2)$ 最优解 $\hat{x}(t)$.

(d) 还设问题 $(\mathcal{DQ})$ 具有唯一的最优解 $\hat{h}$, 令 $\bar{x}(t)$ 是 $(\mathrm{P}_{u(t)})$ 的 $o(t^2)$ 最优的且 Lipschitz 稳定的解. 则

$$\bar{x}(t) = x_0 + t\hat{h} + o(t). \tag{4.245}$$

证明 因为对 $\mathcal{A}(h, d) := \mathcal{T}_K^2(h, d)$, 极小 – 极大问题 (4.242) 与问题 $(\mathcal{DQ})$ 重合, 由定理 4.85 与定理 4.100 有

$$\lim_{t \downarrow 0} \frac{v(u(t)) - v(u_0) - t\ \mathrm{val}(\mathrm{PL}_d)}{\frac{1}{2}t^2} = \mathrm{val}(\mathcal{DQ}). \tag{4.246}$$

设 h 是 $(\bar{x}(t) - x_0)/t$ 的聚点. 因为 X 是有限维的且 $\bar{x}(t)$ 是 Lipschitz 稳定的, 这样的一个聚点是存在的. 由定理 4.100 证明的最后一不等式, 对任何 $\lambda \in \mathcal{S}(\mathrm{DL}_d)$, 有 (4.246) 左端的值均不小于 $\theta_h(\lambda)$. 从而有 $\mathrm{val}(\mathcal{DQ}) > -\infty$, 还有 h 是 (PL_d) 的最优解, 因而 $\mathcal{S}(\mathrm{PL}_d)$ 非空. 因为 $\mathcal{S}(\mathrm{DL}_d)$ 是有界的, 所以 $\mathrm{val}(\mathcal{DQ}) < +\infty$. 得到 $\mathrm{val}(\mathcal{DQ})$ 是有限的, h 是 $(\mathcal{DQ})$ 的最优解. 证得结论 (a) 与 (b).

由方向正则性条件, 存在由 (c) 得出的 $(P_{u(t)})$ 的可行路径 $\hat{x}(t)$. 进一步, $f(\hat{x}(t), u(t)) = v(u(t)) + o(t^2)$, 从而结论 (c) 成立. 结论 (d) 是 (b) 的一个显然的结果. □

现在已经讨论了最优值函数沿单个 (固定的) 扰动路径的展开式. 通过用更强的条件, 可以得到最优值函数沿方向 d Hadamard 意义下的二阶方向可微性, 即对任何 $r \in U$,

$$\lim_{\substack{t\downarrow 0\\ r'\to r}} \frac{v\left(u_0 + td + \frac{1}{2}t^2 r'\right) - v(u_0) - t\, v'(u_0, d)}{\frac{1}{2}t^2} = v''(u_0; d, r). \tag{4.247}$$

定理 4.102 设

(i) 空间 X 是有限维的.

(ii) x_0 是 (P_{u_0}) 的唯一最优解.

(iii) 沿方向 d 的方向正则性条件成立.

(iv) 方向二阶充分条件 (4.138) 在 x_0 处成立.

(v) 对任何形式为 $\hat{u}_t = u_0 + td + \frac{1}{2}t^2 r + o(t^2)$, 其中 $t > 0$ 充分小的路径, 存在一收敛到 x_0 的问题 $(\mathrm{P}_{\hat{u}_t})$ 的 $o(t^2)$ 最优解路径 $\bar{x}(t)$.

则最优值函数在 u_0 处沿方向 d 是二阶 Hadamard 方向可微的, 且

$$v'(u_0, d) = \mathrm{val}(\mathrm{PL}_d), \quad v''(u_0; d, r) = \mathrm{val}(\mathcal{DQ}). \tag{4.248}$$

进一步, $(\mathrm{P}_{u(t)})$ 的任何 $o(t^2)$ 最优解 $\bar{x}(t)$ 在 x_0 处均是 Lipschitz 稳定的, 且 $t \downarrow 0$ 时的 $(\bar{x}(t) - x_0)/t$ 的每一聚点是 $(\mathcal{DQ})$ 的最优解.

证明 由 (v) 与定理 4.55, 得到 $\bar{x}(t) - x_0 = O(t)$. 由定理 4.101 有展式 (4.244) 成立且当 $t \downarrow 0$ 时 $(\bar{x}(t) - x_0)/t$ 的每一聚点均是 $(\mathcal{DQ})$ 的最优解. 结论成立. □

4.7.4 复合最优化问题

这一节考虑如下的参数化复合最优化问题

$$(\mathrm{P}_u) \qquad \min_{x\in X} g(F(x,u)), \tag{4.249}$$

其中 $g : Y \to \overline{\mathbb{R}}$ 是下半连续的正常的凸函数, $F : X \times U \to Y$, X, Y 与 U 是 Banach 空间. 3.4.1 节讨论了这样问题的一阶与二阶最优性条件. 如 3.4.1 节指出的那样, 上述复合最优化问题可以表示为下述等价形式

$$\min_{(x,c)\in X\times\mathbb{R}} c \quad \text{s.t.} \quad (F(x,u), c) \in \mathrm{epi}\, g. \tag{4.250}$$

在上述形式中, 这一问题成为参数化问题 (4.1) 的特殊形式, 因此前面几节所发展的扰动分析方法可以直接应用. 这一节将把前面得到的结果转换成复合最优化的语言.

设映射 $F(x,u)$ 是连续可微的. 可以把问题 (4.250) 在点 (x_0,u_0) 处沿给定方向 $d\in U$ 相应的线性化问题表示为下述形式

$$\min_{(h,c)\in X\times\mathbb{R}} c \quad \text{s.t.} \quad (DF(x_0,u_0)(h,d),c)\in T_{\text{epi } g}(y_0,c_0), \tag{4.251}$$

其中 $c_0:=g(y_0)$, $y_0:=F(x_0,u_0)$. 用公式 (2.100) 计算集合 epi g 的切锥, 上述问题可以表示为下述等价形式

$$(\text{PL}_d) \qquad \min_{h\in X} g_-^{\downarrow}(y_0,DF(x_0,u_0)(h,d)). \tag{4.252}$$

上述问题 (PL_d) 的对偶问题是

$$(\text{DL}_d) \qquad \max_{\lambda\in\Lambda(x_0,u_0)} \langle\lambda,D_uF(x_0,u_0)d\rangle, \tag{4.253}$$

其中 $\Lambda(x_0,u_0)$ 是满足非扰动问题一阶最优性条件 (3.207) 的 Lagrange 乘子的集合. 这一对偶问题可用复合问题的一般对偶形式 (3.203) 或者类似于 (4.46) 的形式推导出来.

我们有最优值函数方向导数的下述公式

$$v'(u_0,d)=\inf_{x\in\mathcal{S}(u_0)}\sup_{\lambda\in\Lambda(x_0,u_0)}\langle\lambda,D_uF(x_0,u_0)d\rangle. \tag{4.254}$$

由 4.2.3 节的结果, 这一公式在方向正则条件下, 当问题 (P_{u_0}) 是定义 2.161 意义下的凸问题时成立, 或当确保最优解的 Lipschitz 稳定性的充分的强二阶条件时这一公式也是成立的.

现在设 $F(x,u)$ 是二阶连续可微的. 考虑具有下述形式的路径

$$u(t):=u_0+td+\frac{1}{2}t^2r+o(t^2), \quad t\geqslant 0 \tag{4.255}$$

与极小 - 极大问题

$$(\mathcal{DQ}) \qquad \min_{h\in\mathcal{S}(\text{PL}_d)}\max_{\lambda\in\mathcal{S}(\text{DL}_d)}\xi_h(\lambda), \tag{4.256}$$

其中

$$\xi_h(\lambda):=\langle\lambda,D_uF(x_0,u_0)r+D^2F(x_0,u_0)((h,d),(h,d))\rangle-\psi_h^*(\lambda), \tag{4.257}$$

$$\psi_h(\cdot):=g_-^{\downarrow\downarrow}(y_0;DF(x_0,u_0)(h,d),\cdot), \tag{4.258}$$

ψ_h^* 是 ψ_h 的共轭. 在复合最优化中, sigma 项取函数 ψ_h 的共轭形式 (见 (3.217)). 因此, 上述问题 $(\mathcal{DQ})$ 是定义在 (4.216) 中的相对应的问题 $(\mathcal{DQ})$ 在当前情形下的特殊形式.

回顾凸函数的二阶正则性的定义 3.93. 下述定理是定理 4.101 的一个重新表示.

定理 4.103 设

(i) 空间 X 是有限维的.

(ii) $\mathcal{S}(u_0)=\{x_0\}$, $\Lambda(x_0,u_0)\neq\varnothing$.

(iii) 方向正则性条件成立.

(iv) 存在 $(\mathrm{P}_{u(t)})$ 的 $o(t^2)$ 最优解 $\bar{x}(t)$, 它在 x_0 处是 Lipschitz 稳定的.

(v) 函数 g 在点 $y_0:=F(x_0,u_0)$ 处是二阶正则的.

则下述结论成立:

(a) 对 $t\geqslant 0$,

$$v(u(t))=v(u_0)+t\ \mathrm{val}(\mathrm{PL}_d)+\frac{t^2}{2}\mathrm{val}(\mathcal{DQ})+o(t^2). \tag{4.259}$$

(b) 对 $(\mathrm{P}_{u(t)})$ 的每一 $o(t^2)$ 最优解 $\bar{x}(t)$, 当 $t\downarrow 0$ 时 $(\bar{x}(t)-x_0)/t$ 的每一聚点均是 $(\mathcal{DQ})$ 的最优解.

(c) 进一步, 设问题 $(\mathcal{DQ})$ 具有唯一最优解 $\hat{h}$, 令 $\bar{x}(t)$ 是 $(\mathrm{P}_{u(t)})$ 的 $o(t^2)$ 最优解, 且是 Lipschitz 稳定的. 则

$$\bar{x}(t)=x_0+t\hat{h}+o(t). \tag{4.260}$$

注意到, 为了确保 Lipschitz 稳定的 $o(t^2)$ 最优解 $\bar{x}(t)$ 的存在性, 上述定理中二阶充分性条件的一个强形式被隐式地用到. 也注意到, 若 epi g 是多面体集合, 则问题 $(\mathcal{DQ})$ 中的项 $\psi_h^*(\lambda)$ 消失, 此种情形, 4.7.2 节中的结果是可应用的.

例 4.104 考虑参数化问题

$$(\mathrm{P}_u) \qquad \min_{x\in X}\{\max_{1\leqslant i\leqslant m} f_i(x,u)\}. \tag{4.261}$$

通过定义 $g:\mathbb{R}^m\to\mathbb{R}$, $F:X\times U\to\mathbb{R}^m$,

$$g(y):=\max_{1\leqslant i\leqslant m}(y_i),\quad F(x,u):=(f_1(x,u),\cdots,f_m(x,u)),$$

可以在复合最优化的框架下考虑问题 (4.261). 注意到最大值函数 g 的上图是多面体, 因此相应的 sigma 项消失, 且

$$g'(y,z)=\max_{i\in I(y)} z_i,$$

其中 $I(y):=\{i:y_i=g(y),i=1,\cdots,m\}$.

该函数 $f_i(x,u)$ 是二次连续可微的. (4.261) 的线性化问题是

$$(\mathrm{PL}_d) \qquad \min_{h\in X}\{\max_{i\in I(y_0)} Df_i(x_0,u_0)(h,d)\}, \tag{4.262}$$

其中 $y_0 := (f_1(x_0, u_0), \cdots, f_m(x_0, u_0))$, 集合 $\Lambda(x_0, u_0)$ 由满足下述一阶最优性条件的 Lagrange 乘子 $\lambda = (\lambda_1, \cdots, \lambda_m)$ 构成

$$\sum_{i=1}^{m} \lambda_i D_x f_i(x_0, u_0) = 0, \quad \sum_{i=1}^{m} \lambda_i = 1, \quad \lambda \geqslant 0, \ \lambda_i = 0, \ i \neq I(y_0). \tag{4.263}$$

注意到, 在此种情形中, 函数 g 是连续的, 因此集合 $\Lambda(x_0, u_0)$ 是非空的有界的.

问题 (PL_d) 的对偶问题是

$$(\mathrm{DL}_d) \qquad \max_{\lambda \in \Lambda(x_0, u_0)} \sum_{i=1}^{m} \lambda_i D_u f_i(x_0, u_0) d. \tag{4.264}$$

相应的问题 $(\mathcal{DQ})$ 具有下述形式

$$\min_{h \in \mathcal{S}(\mathrm{PL}_d)} \max_{\lambda \in \mathcal{S}(\mathrm{DL}_d)} \left\{ \sum_{i=1}^{m} \lambda_i [D_u f_i(x_0, u_0) r + D^2 f_i(x_0, u_0)((h, d), (h, d))] \right\}. \tag{4.265}$$

由定理 4.94 与 4.95 得, 若 X 是自反的 Banach 空间, $(\mathrm{P}_{u(t)})$ 具有在 x_0 处 Lipschitz 稳定的 $o(t^2)$ 最优解 $\bar{x}(t)$, 函数 $D_{xx}^2 f_i(x_0, u_0)(\cdot, \cdot)$ 是弱下半连续的, 则

$$\lim_{t \downarrow 0} \frac{v(u(t)) - v(u_0) - t \ \mathrm{val}(\mathrm{PL}_d)}{\frac{1}{2} t^2} = \mathrm{val}(\mathcal{DQ}). \tag{4.266}$$

若还有, 对集合 $\mathcal{S}(\mathrm{DL}_d)$ 的任何 λ, $\sum_{i=1}^{m} \lambda_i D_{xx}^2 f_i(x_0, u_0)(\cdot, \cdot)$ 是 Legendre 形式, 且 $(\mathcal{DQ})$ 有唯一的最优解 $\bar{h}$, 则

$$x(t) = x_0 + t \ \bar{h} + o(t), \quad t \geqslant 0. \tag{4.267}$$

若空间 X 是有限维的, 则上述弱下半连续性与 Legendre 形式的条件自动成立.

例 4.105 现在考虑参数化问题

$$(\mathrm{P}_u) \qquad \min_{x \in X} \{ f(x) + \varphi(x, u) \}, \tag{4.268}$$

其中 $f : X \to \overline{\mathbb{R}}$ 是正常的下半连续的凸函数, $\varphi(x, u)$ 是二次连续可微函数. 定义

$$g(x, c) := f(x) + c, \quad F(x, u) := (x, \varphi(x, u)),$$

我们以复合最优化问题 (4.249) 的框架考虑问题 (4.268). 有

$$\mathrm{dom} \ g = \mathrm{dom} \ f \times \mathbb{R}, \quad D_x F(x_0, u_0) h = (h, D_x \varphi(x_0, u_0) h).$$

因此, 对上述最优化问题, Robinson 约束规范 (3.202) 总是成立的. 存在唯一的 Lagrange 乘子

$$\lambda = (-D_x\phi(x_0, u_0), 1)$$

满足一阶最优性条件 (3.207). 注意, 这里的一阶最优性条件可以写为下述形式

$$0 \in \partial f(x_0) + D_x\varphi(x_0, u_0), \tag{4.269}$$

或由命题 2.126(ii), 等价地表示为

$$f_-^\downarrow(x_0, h) + D_x\varphi(x_0, u_0)h \geqslant 0, \quad \forall h \in X. \tag{4.270}$$

相应的线性化问题具有下述形式

$$(\mathrm{PL}_d) \qquad \min_{h\in X} f_-^\downarrow(x_0, h) + D_x\varphi(x_0, u_0)(h, d). \tag{4.271}$$

由一阶最优性条件 (4.270), 上述问题 (PL_d) 的最优解集与临界方向锥重合

$$C(x_0) = \{h \in X : f_-^\downarrow(x_0, h) + D_x\varphi(x_0, u_0)h = 0\}, \tag{4.272}$$

由命题 2.126(ii), 也可以表示为

$$C(x_0) = N_{\partial f(x_0)}(-D_x\varphi(x_0, u_0)). \tag{4.273}$$

由 (4.258) 定义的函数 ψ_h 成为 $\psi_h(w, c) = \phi_h(w) + c$, 其中

$$\phi_h(\cdot) := f_-^{\downarrow\downarrow}(x_0; h, \cdot). \tag{4.274}$$

结果相对应的问题 ($\mathcal{DQ}$) 取下述形式

$$D_u\varphi(x_0, u_0)r + \min_{h\in C(x_0)}\{D^2\varphi(x_0, u_0)((h, d), (h, d)) - \phi_h^*(-D_x\varphi(x_0, u_0))\}^{①} \tag{4.275}$$

例 4.106 设 X 是有限维空间, 如 $X := \mathbb{R}^n$, 赋予范数 $\|\cdot\|$ 为欧氏范数, 令 $f : X \to \overline{\mathbb{R}}$ 是正常的下半连续的凸函数. 考虑 f 的 Moreou-Yosida 正则化, 即

$$\hat{f}_\varepsilon(u) := \min_{x\in X}\left\{f(x) + \frac{1}{2}\varepsilon\|x - u\|^2\right\}, \quad \varepsilon > 0. \tag{4.276}$$

在例 4.33 中已经讨论了 Moreou-Yosida 正则化的某些基本性质. 若定义 $\varphi(x, u) := \frac{1}{2}\varepsilon\|x - u\|^2$, 则上述问题是问题 (4.268) 的特殊形式.

① 原著中 $C(x_0)$ 被写为 C_0.

考虑点 u_0, 令 $x_0 := \bar{x}(u_0)$ 是 (4.276) 的最优解. 有

$$D\varphi(x_0, u_0)(h, d) = \varepsilon\langle x_0 - u_0, h - d\rangle, \tag{4.277}$$

$$D^2\varphi(x_0, u_0)((h, d), (h, d)) = \varepsilon\|h - d\|^2. \tag{4.278}$$

结果, 一阶最优性条件可以写为

$$\varepsilon(u_0 - x_0) \in \partial f(x_0), \tag{4.279}$$

临界锥 $C(x_0)$ 为

$$\begin{aligned} C(x_0) &= \{h : f_-^{\downarrow}(x_0, h) + \varepsilon\langle x_0 - u_0, h\rangle = 0\} \\ &= N_{\partial f(x_0)}(\varepsilon(u_0 - x_0)), \end{aligned} \tag{4.280}$$

从而问题 $(\mathcal{DQ})$ 取下述形式

$$(\mathcal{DQ}) \qquad \varepsilon\langle u_0 - x_0, r\rangle + \min_{h \in C(x_0)} \left\{\varepsilon\|h - d\|^2 - \phi_h^*(\varepsilon(u_0 - x_0))\right\}, \tag{4.281}$$

其中 $\phi_h(\cdot) := f_-^{\downarrow\downarrow}(x_0; h, \cdot)$ 且

$$\phi_h^*(\varepsilon(u_0 - x_0)) = \sup_w \left\{\varepsilon\langle u_0 - x_0, w\rangle - f_-^{\downarrow\downarrow}(x_0; h, w)\right\}.$$

我们得到, 若 f 在 x_0 处是二阶正则的, 则 $\hat{f}_\varepsilon''(u_0; d, r)$ 存在且等于问题 (4.281) 的最优值. 进一步, $\bar{x}(u)$ 是方向可微的, 因为 $\bar{x}(\cdot)$ 在 u_0 处是 Lipschitz 连续的, 因而是 Hadamard 方向可微的, 且 $\bar{x}'(u_0, d) = \hat{h}$, 其中 $\hat{h}$ 是 (4.281) 的最优解.

例如, 设

$$f(\cdot) := \max_{1 \leqslant i \leqslant m} f_i(\cdot),$$

其中 $f_i, i = 1, \cdots, m$ 是二次连续可微凸函数. 此种情形, 极大值函数 f 是二阶正则的和二阶方向可微的, 且 (见例 2.68)

$$f'(x_0, h) = \max_{i \in I(x_0)} Df_i(x_0)h.$$

进一步

$$\phi_h(w) := f''(x_0; h, w) = \max_{i \in I_1(x_0, h)} \left\{Df_i(x_0)w + D^2 f_i(x_0)(h, h)\right\},$$

其中

$$I_1(x_0, h) := \{i : Df_i(x_0)h = f'(x_0, h),\ i = 1, \cdots, m\}$$

于是有[①]

$$\partial f(x_0)=\operatorname{conv}\{\bigcup_{i\in I(x_0)} Df_i(x_0)\},$$

因此, 由一阶最优性条件, 存在乘子 α_i 满足

$$\alpha_i\geqslant 0,\ \sum_{i\in I(x_0)}\alpha_i=1, \varepsilon(u_0-x_0)=\sum_{i\in I(x_0)}\alpha_i Df_i(x_0).$$

用 A_0 记满足上述条件的乘子 $\alpha=(\alpha_i)_{i\in I(x_0)}$ 集合.

得到

$$C(x_0)=\left\{h\in X:\max_{i\in I(x_0)} Df_i(x_0)h+\varepsilon\langle x_0-u_0,h\rangle=0\right\}.$$

则对 $h\in C(x_0)$ 有

$$\begin{aligned}\phi_h^*(\varepsilon(u_0-x_0))&=\sup_{w\in X}\left\{\varepsilon\langle u_0-x_0,w\rangle-\max_{i\in I(x_0)}\{Df_i(x_0)w+D^2f_i(x_0)(h,h)\}\right\}\\&=-\max_{\alpha\in A_0}\sum\alpha_i D^2f_i(x_0)(h,h).\end{aligned}$$

结果, 相应的问题 $(\mathcal{DQ})$ 可以表示为下述形式

$$\varepsilon\langle u_0-x_0,r\rangle+\min_{h\in C(x_0)}\left\{\varepsilon\|h-d\|^2+\max_{\alpha\in A_0}\sum\alpha_i D^2 f_i(x_0)(h,h)\right\}.$$

尤其, 若 $A_0=\{\alpha_0\}$ 是单点集, α_0 的所有分量均是正的, 则 $\varepsilon(u_0-x_0)$ 属于 $\partial f(x_0)$ 的相对内部, 因而 $C(x_0)$ 是线性空间. 此种情形, 问题 $(\mathcal{DQ})$ 变为在线性空间上的二次函数的极小化问题, 因此其最优解是关于 d 为线性的, 最优值关于 d 是二次的. 此种情形, $\bar{x}(u)$ 是可微的, $\hat{f}_\varepsilon(u)$ 在 u_0 处是二阶可微的.

例 4.107 设 $f:X\to\overline{\mathbb{R}}$ 是正常的下半连续凸函数, 考虑问题

$$(\mathrm{P}_x)\qquad \min_{x^*\in X^*}\{f^*(x^*)-\langle x^*,x\rangle\}.\tag{4.282}$$

由 2.4.2 节可以看到, 次微分$\partial f(x)$ 构成了 (P_x) 的最优解的集合, 由 Fenchel-Moreau-Rockafellar 定理(定理 2.113), $f(x)=-\mathrm{val}(\mathrm{P}_x)$(与例 4.34 相比较). 显然, 问题 (P_x) 可以在问题 (2.268) 的框架下考虑.

设空间 X 是有限维的, 共轭函数f^* 是强凸的. 有 $\partial f(x)=\{\nabla f(x)\}$ 是单点集, 进一步, $\nabla f(\cdot)$ 在 X 上是 Lipschitz 连续的 (见例 4.34). 令 $x_0\in X$ 是给定点, $x_0^*:=\nabla f(x_0)$, 即 x_0^* 是问题 (P_{x_0}) 的最优解.

① 上述定义中应为 $i\in I(x_0)$.

具体到当前的问题, 有

$$C(x_0^*) = N_{\partial f^*(x_0^*)}(x_0), \tag{4.283}$$

相应的问题 $(\mathcal{DQ})$ 变成

$$(\mathcal{DQ}) \qquad -\langle x_0^*, r\rangle - \max_{h^* \in C(x_0^*)} \{\langle h^*, d\rangle + \phi_{h^*}^*(0)\}, \tag{4.284}$$

其中 $\phi_{h^*}(\cdot) := f^{*\prime\prime}(x_0^*; h^*, \cdot)$. 我们得到, 若 f^* 在 x^* 处是二阶正则的, 则 $f''(x_0; d, r)$ 存在且

$$f''(x_0; d, r) = -\text{val}(\mathcal{DQ}).$$

进一步有, $\nabla f(x)$ 在 x_0 处是 Hadamard 方向可微的, 且 $(\nabla f)'(x_0, d) = \hat{h}^*$, 其中 $\hat{h}^*$ 是由 (4.284) 给出的问题 $(\mathcal{DQ})$ 的最优解.

4.8 Hölder 稳定性情形的二阶分析

这一节研究 (近似) 最优解关于参数 u 的方向扰动是 $\frac{1}{2}$ 度 Hölder 稳定的情况. 这样的 Hölder 稳定性在方向正则性条件与标准的二阶充分条件之下可以得到, 见定理 4.53. 这一节考虑下述形式的扰动

$$u(t) := u_0 + td + o(t), \quad t \geqslant 0, \tag{4.285}$$

其中 $d \in U$ 是某一 (固定的) 方向. 通过导出沿路径 $u(t)$ 的最优值函数的上方近似开始我们的分析, 则如在 Lipschitz 稳定解的情形, 证得在两种独立的情形, 即在方向的广义多面性条件及二阶正则性成立的条件之下, 这一上方近似是精确的. 我们另外讨论 Lagrange 乘子集合是空集的情况.

4.8.1 最优值函数的上二阶近似

这一节导出最优值函数的上二阶近似. 令 x_0 是非扰动问题的最优解, 令 $h \in C(x_0)$ 是临界方向. 考虑内二阶切集

$$\mathcal{T}_K^2(h) := T_K^{i,2}(G(x_0, u_0), D_x G(x_0, u_0)h). \tag{4.286}$$

这一节导出的上近似基于下述形式的路径

$$x(t) := x_0 + t^{1/2}h + tw + o(t). \tag{4.287}$$

由 $G(\cdot,\cdot)$ 在 (x_0, u_0) 的二阶 Taylor 展式有

$$\begin{aligned} G(x(t), u(t)) = {} & G(x_0, u_0) + t^{1/2} D_x G(x_0, u_0)h \\ & + \frac{1}{2}t[2DG(x_0, u_0)(w, d) + D_{xx}^2 G(x_0, u_0)(h, h)] + o(t). \end{aligned} \tag{4.288}$$

由内二阶切集的定义, 若这样的路径对充分小的 $t>0$ 是可行的, 则

$$2DG(x_0,u_0)(w,d)+D_{xx}^2G(x_0,u_0)(h,h)\in \mathcal{T}_K^2(h). \tag{4.289}$$

以类似于处理 Lipschitz 稳定解情形的方式开始讨论, 尽管在分析中一些具体的技术性困难会出现. 包含关系 (4.289) 启示我们, 为导出最优值函数方向变化的精确的上方近似, 应该在满足约束 (4.289) 的前提下极小化目标函数展式的第二项 (即一阶项) 来选取 w, 即选取 w 为下述问题的最优解

$$(\mathrm{PQ}_h^2)\qquad \begin{array}{ll}\min_w & 2Df(x_0,u_0)(w,d)+D_{xx}^2f(x_0,u_0)(h,h),\\ \text{s. t.} & 2DG(x_0,u_0)(w,d)+D_{xx}^2G(x_0,u_0)(h,h)\in \mathcal{T}_K^2(h).\end{array}$$

向量 h 的最好的选取是取下述问题的最优解

$$(\mathcal{PQ}^2)\qquad \min_{h\in C(x_0)} \operatorname{val}(\mathrm{PQ}_h^2) \tag{4.290}$$

(上标 2 指的是我们在讨论第二形式的展式这一事实). 得到结果

命题 4.108 设 $x_0\in\mathcal{S}(u_0)$, 与之相联系的非空的 Lagrange 乘子集合为 $\Lambda(x_0,u_0)$, 在 x_0 处沿方向 d 的方向正则性成立. 则

$$\limsup_{t\downarrow 0}\frac{v(u(t))-v(u_0)}{t}\leqslant \frac{1}{2}\operatorname{val}(\mathcal{PQ}^2)\leqslant \operatorname{val}(\mathrm{PL}_d). \tag{4.291}$$

注意到, 在方向正则性的前提下, 当 Lagrange 乘子集 $\Lambda(x_0,u_0)$ 非空时, $\operatorname{val}(\mathrm{PL}_d)$ 是有限的 (见命题 4.21). 则由 (4.291) 的第二不等式可得 $\operatorname{val}(\mathcal{PQ}^2)<+\infty$.

上述命题 4.108 的证明基于下述结果.

引理 4.109 设方向正则性条件在可行点 $x_0\in\Phi(u_0)$ 沿方向 d 成立, 令 $u(t):=u_0+td+o(t)$, $x(\cdot):\mathbb{R}_+\to X$ 是当 $t\downarrow 0$ 满足 $x(t)\to x_0$ 的路径且

$$\operatorname{dist}(G(x(t),u(t)),K)=o(t),\quad t\geqslant 0. \tag{4.292}$$

则对充分小的 $\theta>0$, 存在另外一路径 $\bar{x}(\cdot):\mathbb{R}_+\to X$ 满足对 $t\geqslant 0$ 充分小,

$$G(\bar{x}(t),u((1+\theta)t))\in K, \tag{4.293}$$

$$\|\bar{x}(t)-x(t)\|=c\theta t, \tag{4.294}$$

其中常数 c 不依赖于 θ.

证明 令 $y(t)\in K$ 满足

$$\|y(t)-G(x(t),u(t))\|\leqslant 2\operatorname{dist}(G(x(t),u(t)),K). \tag{4.295}$$

因此, 由 (4.292) 得

$$y(t) = G(x(t), u(t)) + o(t),$$

因此, 由 $G(\cdot,\cdot)$ 的连续性, 当 $t \downarrow 0$ 时, $y(t) \to G(x_0, u_0)$. 由定理 4.9 中的包含关系 (4.13), 方向正则性可推出, 对充分小的 $t > 0$, 有

$$\varepsilon' B_Y \subset y(t) + DG(x_0, u_0)(B_X, \delta d) - K \tag{4.296}$$

对某一 $\varepsilon' > 0$ 成立. 令 $\theta \in (0,1), t \in (0, \theta^{-1}\delta)$. 取 (4.296) 包含关系的权值 $t\delta^{-1}\theta$ 与包含关系 $0 \in y(t) - K$ 的权值 $1 - t\delta^{-1}\theta$ 的凸组合, 得到

$$t\theta\varepsilon B_Y \subset y(t) + t\theta DG(x_0, u_0)(\delta^{-1} B_X, d) - K, \tag{4.297}$$

其中 $\varepsilon := \varepsilon'/\delta$. 尤其, 存在 $z_t \in \delta^{-1} B_X$ 满足

$$0 \in y(t) + t\theta DG(x_0, u_0)(z_t, d) - K. \tag{4.298}$$

由于

$$\begin{aligned} y(t) + t\theta DG(x_0, u_0)(z_t, d) &= G(x(t), u(t)) + t\theta DG(x_0, u_0)(z_t, d) + o(t) \\ &= G(x(t) + t\theta z_t, u((1+\theta)t)) + o(t), \end{aligned}$$

由 (4.297) 与 (4.298), 对 $t > 0$ 充分小, 得到 $y'(t) := y(t) + t\theta DG(x_0, u_0)(z_t, d)$ 满足 $y'(t) \in K$, 且

$$\|G(x(t) + t\theta z_t, u((1+\theta)t)) - y'(t)\| = o(t), \tag{4.299}$$

$$t\theta\varepsilon B_Y \subset y'(t) + 2t\delta^{-1}\theta D_x G(x_0, u_0) B_X - K. \tag{4.300}$$

考虑多值函数

$$\Psi(x) := y'(t) + D_x G(x_0, u_0)(x - x(t) - t\theta z_t) - K.$$

由 (4.300) 得到

$$t\theta\varepsilon B_Y \subset \Psi(x(t) + t\theta z_t + 2t\delta^{-1}\theta B_X).$$

因为 Ψ 是凸的, 由命题 2.77 知多值函数 Ψ 在 $(x(t) + t\theta z_t, 0)$ 处是开的, 在包含关系 (2.130) 中相应的常数为

$$\eta := t\theta\varepsilon, \quad v := 2t\delta^{-1}\theta.$$

结果, 由定理 2.83, 只要 $(x, y) \in N := \eta_x B_X \times \eta_y B_Y$, 以 c 为率的度量正则性条件成立, 其中

$$\eta_x := \frac{1}{2}v = t\delta^{-1}\theta, \quad \eta_y := \eta/8 = t\theta\varepsilon/8, \quad c := 4v/\eta = 8\varepsilon^{-1}\delta^{-1}.$$

现在, 应用 (4.18) 和稳定性定理 (定理 2.84) 推导 (4.294) 中的估计. 为避免混淆, 这里用 G' 与 H' 记此定理中的 G 与 H, 即

$$\begin{aligned} G'(x) &:= y'(t) + D_xG(x_0, u_0)(x - x(t) - t\theta z_t), \\ H'(x) &:= G(x, u((1+\theta)t)). \end{aligned}$$

该定理中的点 x_0 与 y_0 分别是 $x(t) + t\theta z_t$ 与 $0 \in Y$. 考虑到 $D(x) := G'(x) - H'(x)$, 由 (4.299) 得

$$\|D(x(t) + t\theta z_t)\| = \|y'(t) - G(x(t) + t\theta z_t, u((1+\theta)t)\| = o(t).$$

$D(\cdot)$ 在 x_0 的邻域的 Lipschitz 常数 κ 是有界的, 其上界为 $\|D_xG(x_0, u_0) - D_xG(x, u((1+\theta)t))\|$ 在该邻域的上确界, 对充分小的 $t > 0$, 当靠近 $x(t) + t\theta z_t$ 时, 它是任意小的. 因为可以把 (2.160) 与 (2.161) 中的 η_x^1 与 η_y^1 取为 $o(t)$ 函数, 对充分小的 $t > 0$, 这些不等式是成立的. 取 $x := x(t) + t\theta z_t$ 与 $y = 0$, 由稳定性定理 (定理 2.84) 的估计式 (2.145) 可得

$$\operatorname{dist}(x(t) + t\theta z_t, \Phi(u((1+\theta)t))) \leqslant c(\kappa) \operatorname{dist}(G(x(t) + t\theta z_t, u((1+\theta)t)), K).$$

由 (4.299), 上述不等式的右端是 $o(t)$ 阶的, 则存在 $\bar{x}(t)$ 满足 (4.293), 同时

$$\|\bar{x}(t) - x(t)\| \leqslant t\theta\delta^{-1} + o(t).$$

不等式 (4.294) 成立, 其中 $c := 2\delta^{-1}$. □

命题 4.108 的证明　证明 (4.291) 中的第一不等式. 令 $h \in C(x_0)$, w 是 (PQ_h^2) 的可行点. 由二阶切集的定义, 路径 $x(t) := x_0 + t^{1/2}h + tw$ 满足 (4.292). 由引理 4.109, 对充分小 $\theta > 0$, 存在路径 $\bar{x}(\cdot) : \mathbb{R}_+ \to X$ 满足对充分小的 $t \geqslant 0$, (4.293) 与 (4.294) 成立. 因此

$$\begin{aligned} \limsup_{t \downarrow 0} \frac{v(u(t)) - v(u_0)}{t} &= \limsup_{t \downarrow 0} \frac{v(u((1+\theta)t)) - v(u_0)}{(1+\theta)t} \\ &\leqslant \limsup_{t \downarrow 0} \frac{f(\bar{x}(t), u((1+\theta)t)) - f(x_0, u_0)}{(1+\theta)t} \\ &= (1+\theta)^{-1} Df(x_0, u_0)(w, (1+\theta)d) + O(\theta). \end{aligned}$$

取 $\theta \downarrow 0$ 时的极限, 再取 $w \in F$ 的极小化, 其中 F 是问题 (PQ_h^2) 的可行集, 得到

$$\limsup_{t \downarrow 0} \frac{v(u(t)) - v(u_0)}{t} \leqslant \inf_{w \in F} Df(x_0, u_0)(w, d) \leqslant \operatorname{val}(\mathrm{PQ}_h^2).$$

取关于 $h \in C(x_0)$ 的极小化, 得到 (4.291) 的第一不等式.

现在证明 (4.291) 中的第二不等式. 考虑临界方向 $h=0$. 由 (4.286) 定义的相应的内二阶切集 $\mathcal{T}_K^2(h)$ 与 $T_K(G(x_0,u_0))$ 重合. 因此, 问题 (PQ_0^2) 与线性化问题 (PL_d) 是重合的 (相差 2 倍的量). 因此, 由 $(\mathcal{PQ}^2)$ 的定义, 得到

$$\frac{1}{2}\mathrm{val}(\mathcal{PQ}^2) \leqslant \frac{1}{2}\mathrm{val}(\mathcal{PQ}_0^2) = \mathrm{val}(\mathrm{PL}_d). \qquad \square$$

问题 (PQ_h^2) 关于 w 是线性的, 其对偶可类似于问题 (PQ_h) 的对偶 (DQ_h)(见 (4.208) 后面的推导) 的推导, 即 (PQ_h^2) 的对偶可表示为

$$(\mathrm{DQ}_h^2) \qquad \max_{\lambda\in\Lambda(x_0,u_0)} \xi_h^2(\lambda), \tag{4.301}$$

其中 (考虑由 (4.286) 定义的 $\mathcal{T}_K^2(h)$)

$$\xi_h^2(\lambda) := 2D_uL(x_0,\lambda,u_0)d + D_{xx}^2L(x_0,\lambda,u_0)(h,h) - \sigma(\lambda,\mathcal{T}_K^2(h)). \tag{4.302}$$

考虑问题

$$(\mathcal{DQ}^2) \qquad \min_{h\in C(x_0)}\max_{\lambda\in\Lambda(x_0,u_0)} \xi_h^2(\lambda). \tag{4.303}$$

下述引理表明, 尽管对于某些临界方向 h, (PQ_h^2) 与 (DQ_h^2) 的最优值可能是不同的, 但问题 $(\mathcal{PQ}^2)$ 与 $(\mathcal{DQ}^2)$ 具有相同的最优值.

引理 4.110 设方向正则性条件在 $x_0\in\Phi(u_0)$ 处沿方向 d 成立, Lagrange 乘子集合 $\Lambda(x_0,u_0)$ 非空. 令 $h\in C(x_0)$ 是临界方向. 则下述结论成立:

(i) $\mathrm{val}(\mathrm{DQ}_h^2)\leqslant \mathrm{val}(\mathrm{PQ}_h^2)$.

(ii) 若对某一 $\gamma>1$, 问题 $(\mathrm{PQ}_{\gamma h}^2)$ 是相容的 (即可行集非空), 则 $\mathrm{val}(\mathrm{PQ}_h^2) = \mathrm{val}(\mathrm{DQ}_h^2)$ 有限, 且集合 $\mathcal{S}(\mathrm{DQ}_h^2)$ 非空有界.

(iii) 若 (PQ_h^2) 是不相容的, 则对所有 $\gamma>1$,

$$\mathrm{val}(\mathrm{DQ}_{\gamma h}^2) = \mathrm{val}(\mathrm{PQ}_{\gamma h}^2) = +\infty.$$

(iv) 下述不等式成立

$$\limsup_{\eta\uparrow 1}[\mathrm{val}(\mathrm{DQ}_{\eta h}^2)] \leqslant \mathrm{val}(\mathrm{DQ}_h^2). \tag{4.304}$$

(v) $\mathrm{val}(\mathcal{DQ}^2) = \mathrm{val}(\mathcal{PQ}^2)$.

证明 (i) 由对偶问题的最优值小于或等于原始问题的最优值 (见 (2.268)) 得到.

(ii) 令 $\lambda\in\Lambda(x_0,u_0)$. 由命题 4.84 得 $\sigma(\lambda,\mathcal{T}_K^2(h))\leqslant 0$, 因此 $\mathrm{val}(\mathrm{DQ}_h^2)>-\infty$. 注意到命题 2.153 与 2.147, 为证明 (ii) 中的结论, 只需要验证正则性条件 $0\in\mathrm{int}(E)$ 成立, 其中

$$E := D_xG(x_0,u_0)X + 2D_uG(x_0,u_0)d + D_{xx}^2G(x_0,u_0)(h,h) - \mathcal{T}_K^2(h).$$

令 $\gamma > 1$, w 是问题 $(\mathrm{PQ}^2_{\gamma h})$ 的可行点. 则

$$2DG(x_0,u_0)(w,d) + D^2_{xx}G(x_0,u_0)(\gamma h,\gamma h) \in T^2_K(\gamma h).$$

由二阶切集的定义, 有 $\mathcal{T}^2_K(\gamma h) = \gamma^2\mathcal{T}^2_K(h)$. 因此, 在上述包含关系的两边同时除以 γ^2, 置 $\hat{w} := \gamma^{-2}w$, 得到

$$2DG(x_0,u_0)(\hat{w},\gamma^{-2}d) + D^2_{xx}G(x_0,u_0)(h,h) \in \mathcal{T}^2_K(h).$$

注意到 (3.34), 并用到包含关系 $T_K(y_0) \subset T_{T_K(y_0)}(D_xG(x_0,u_0)h)$, 得到

$$2DG(x_0,u_0)(\hat{w},\gamma^{-2}d) + D^2_{xx}G(x_0,u_0)(h,h) + T_K(y_0) \subset \mathcal{T}^2_K(h).$$

因此

$$D_xG(x_0,u_0)X + 2(1-\gamma^2)D_u(x_0,u_0)d - T_K(y_0) \subset E,$$

连同 (4.13) 可导致

$$\begin{aligned}\varepsilon B_Y &\subset G(x_0,u_0) + DG(x_0,u_0)(X,\delta d) - K\\ &\subset D_xG(x_0,u_0)X + \delta D_u(x_0,u_0)d - T_K(y_0) \subset \frac{1}{2}\delta(1-\gamma^{-2})^{-1}E,\end{aligned}$$

因而有 $0 \in \mathrm{int}(E)$. 这证得 (ii).

(iii) 若 $\mathcal{T}^2_K(h) = \varnothing$, 则对所有 $\gamma > 0$, 有 $\mathcal{T}^2_K(\gamma h) = \gamma^2\mathcal{T}^2_K(h) = \varnothing$, 因而有 $\mathrm{val}(\mathrm{PQ}^2_h) = \mathrm{val}(\mathrm{DQ}^2_h) = +\infty$. 设 $\mathcal{T}^2_K(h)$ 含有某一 $z_0 \in Y$, 置 $z_1 := D^2_{xx}G(x_0,u_0)(h,h)$. 问题 (PQ^2_h) 是不相容的当且仅当集合

$$F := DG(x_0,u_0)(X,2d) + z_1 - \mathcal{T}^2_K(h)$$

不含有 0, 集合 F 是凸的且具有非空的内部. 事实上, 由命题 3.34 与定理 4.9 有

$$\begin{aligned}\frac{1}{2}\delta F &\supset DG(x_0,u_0)(X,\delta d) + \frac{1}{2}\delta(z_1-z_0) - T_{T_K(y_0)}(D_xG(x_0,u_0)h)\\ &\supset DG(x_0,u_0)(X,\delta d) + \frac{1}{2}\delta(z_1-z_0) - T_K(y_0)\\ &\supset \varepsilon B_Y + \frac{1}{2}\delta(z_1-z_0)\end{aligned}$$

对某一正数 $\varepsilon > 0$ 成立, 因而 $z_1 - z_0$ 是 F 的内点.

因为 $0 \notin F$, F 是具有非空内部的凸集合, 由第二分离定理 (定理 2.14), 存在非零的 $\mu \in Y^*$, 它分离 F 与 0, 即

$$2\langle \mu, DG(x_0,u_0)(w,d) + D^2_{xx}G(x_0,u_0)(h,h)\rangle \geqslant \sigma(\mu,\mathcal{T}^2_K(h)),\quad \forall w \in Y.$$

关于 w 取极小化, 得到 $D_xG(x_0,u_0)^*\mu=0$ 且

$$\Xi(\mu,h):=\langle\mu,2D_uG(x_0,u_0)d+D_{xx}^2G(x_0,u_0)(h,h)\rangle-\sigma(\mu,\mathcal{T}_K^2(h))\geqslant 0.$$

由命题 3.43, 若 μ 不属于 $N_K(G(x_0,u_0))$, 有 $\sigma(\mu,\mathcal{T}_K^2(h))=+\infty$. 因此, 上述不等式表明 $\mu\in N_K(G(x_0,u_0))$, μ 是奇异的乘子, 因而它是 $\Lambda(x_0,u_0)$ 的一个回收方向. 由方向正则性条件, 线性化问题 (PL_d) 的对偶, 即问题 (DL_d) 具有非空有界的最优解集. 因为 (DL_d) 是在 $\Lambda(x_0,u_0)$ 上极大化 $D_uL(x_0,\lambda,u_0)d$, 这表明 $\langle\mu,D_uG(x_0,u_0)d\rangle<0$. 将此式与上述不等式乘以 γ^2 后相结合, 用到 $\mathcal{T}_K^2(\gamma h)=\gamma^2\mathcal{T}_K^2(h)$, 则对任何 $\gamma>1$,

$$\Xi(\mu,\gamma h)=\gamma^2\Xi(\mu,h)+2(1-\gamma^2)\langle\mu,D_uG(x_0,u_0)d)\rangle>0.$$

给定 $\lambda\in\Lambda(x_0,u_0)$, 用到 $\Xi(\cdot,\gamma h)$ 是正齐次的且凹的这样一个事实及 $\lambda+t\mu\in\Lambda(x_0,u_0)$ 对所有的 $t>0$ 成立, 得到

$$\begin{aligned}\mathrm{val}(\mathrm{DQ}_{\gamma h}^2)&\geqslant 2D_uf(x_0,u_0)d+D_{xx}^2f(x_0,u_0)(h,h)+\Xi(\lambda+t\mu,\gamma h)\\&\geqslant 2D_uf(x_0,u_0)d+D_{xx}^2f(x_0,u_0)(h,h)+\Xi(\lambda,\gamma h)+t\Xi(\mu,\gamma h).\end{aligned}$$

注意到 $\Xi(\lambda,\gamma h)$ 是有限值的, 因为 $\mathcal{T}_K^2(h)$ 非空, 由 (3.109) 有 $\sigma(\lambda,\mathcal{T}_K^2(h))\leqslant 0$. 因为上述不等式对任何 $t>0$ 均成立, 且 $\Xi(\mu,\gamma h)>0$, 有 $\mathrm{val}(\mathrm{DQ}_{\gamma h}^2)=+\infty$. 将这一结论与 (i) 结合, 结论 (iii) 得证.

(iv) 对 $\eta\in(0,1)$, 记 $\Lambda_0:=\Lambda(x_0,u_0)$, 有

$$\begin{aligned}\mathrm{val}(\mathrm{DQ}_{\eta h}^2)&=\sup_{\lambda\in\Lambda_0}\{2D_uL(x_0,\lambda,u_0)d+\eta^2D_{xx}^2L(x_0,\lambda,u_0)(h,h)-\eta^2\sigma(\lambda,\mathcal{T}_K^2(h))\}\\&\leqslant\sup_{\lambda\in\Lambda_0}\{2(1-\eta^2)D_uL(x_0,\lambda,u_0)d\}+\eta^2\mathrm{val}(\mathrm{DQ}_h^2)\\&=2(1-\eta^2)\mathrm{val}(\mathrm{PL}_d)+\eta^2\mathrm{val}(\mathrm{DQ}_h^2).\end{aligned}$$

因为 $\mathrm{val}(\mathrm{PL}_d)<+\infty$, 结论 (iv) 得证.

(v) 由结论 (i) 有 $\mathrm{val}(\mathcal{DQ}^2)\leqslant\mathrm{val}(\mathcal{PQ}^2)$, 由命题 4.108, 有

$$\frac{1}{2}\mathrm{val}(\mathcal{PQ}^2)\leqslant\mathrm{val}(\mathrm{PL}_d)<+\infty.$$

现在令 $h\in C(x_0)$ 满足 $\mathrm{val}(\mathrm{DQ}_h^2)<+\infty$. 则由 (iii), 对所有 $\gamma>1$, 问题 $(\mathrm{PQ}_{\gamma^{-1}h}^2)$ 是可行的, 且 $\mathrm{val}(\mathrm{DQ}_{\gamma^{-1}h}^2)=\mathrm{val}(\mathrm{PQ}_{\gamma^{-1}h}^2)$. 则由 (iv), 取 $\eta:=\gamma^{-1}$,

$$\mathrm{val}(\mathcal{PQ}^2)\leqslant\limsup_{\eta\uparrow 1}[\mathrm{val}(\mathrm{DQ}_{\eta h}^2)]\leqslant\mathrm{val}(\mathrm{DQ}_h^2).$$

因为对任何使 (DQ_h^2) 相容的临界方向 h 均成立, 有 $\mathrm{val}(\mathcal{PQ}^2)\leqslant\mathrm{val}(\mathcal{DQ}^2)$. 结论 (v) 得证. □

现在讨论若 sigma 项被移走且 $(\mathcal{P}Q^2)$ 与 $(\mathcal{D}Q^2)$ 的公共最优值不发生变化的情形. 考虑函数

$$\xi_h^{2,R}(\lambda) := 2D_u L(x_0,\lambda,u_0)d + D_{xx}^2 L(x_0,\lambda,u_0)(h,h). \tag{4.305}$$

除了 sigma 项被移走外, 此函数与函数 $\xi_h^2(\lambda)$ 是相同的. 考虑松弛问题

$$(\mathrm{DQ}_h^{2,R}) \qquad \max_{\lambda\in\Lambda(x_0,u_0)} \xi_h^{2,R}(\lambda), \tag{4.306}$$

$$(\mathcal{D}Q^{2,R}) \qquad \min_{h\in C(x_0)} \max_{\lambda\in\Lambda(x_0,u_0)} \xi_h^{2,R}(\lambda). \tag{4.307}$$

若 $\Lambda(x_0,u_0)$ 非空, 则在方向正则性条件之下, 集合 $\mathcal{S}(\mathrm{DL}_d)$ 是非空的弱 * 紧致的. 回顾广义多面性条件的 3.52 定义, 考虑下述二阶条件

$$\sup_{\lambda\in\mathcal{S}(\mathrm{DL}_d)} D_{xx}^2 L(x_0,\lambda,u_0)(h,h) \geqslant 0, \quad \forall h\in C(x_0). \tag{4.308}$$

可将这些条件与二阶条件 (4.138) 相比较.

命题 4.111 设方向正则性条件在 $x_0\in\Phi(u_0)$ 沿方向 d 成立, Lagrange 乘子集合 $\Lambda(x_0,u_0)$ 非空, 则

(i) 若二阶条件 (4.308) 成立, 则

$$\mathrm{val}(\mathcal{D}Q^2) = \mathrm{val}(\mathcal{P}Q^2) = 2\mathrm{val}(\mathrm{PL}_d). \tag{4.309}$$

(ii) 等式

$$\mathrm{val}(\mathcal{D}Q^2) = \mathrm{val}(\mathcal{P}Q^2) = \mathrm{val}(\mathcal{D}Q^{2,R}) \tag{4.310}$$

在下述两个条件之一被满足时成立:

(a) 对每一临界方向 $h\in C(x_0)$, $0\in\mathcal{T}_K^2(h)$;

(b) Robinson 约束规范 (2.163) 与广义的多面性条件成立.

(iii) 若二阶充分条件 (3.136) 在 $x_0\in\Phi(u_0)$ 处成立, 则 $\mathrm{val}(\mathcal{D}Q^{2,R}) > -\infty$.

证明 (i) 令 $h\in C(x_0)$. 因为对所有的 $\lambda\in\Lambda(x_0,u_0)$, $\sigma(\lambda,\mathcal{T}_K^2(h))\leqslant 0$, 得到

$$\mathrm{val}(\mathrm{DQ}_h^2) \geqslant \sup_{\lambda\in\Lambda(x_0,u_0)} \xi_h^{2,R}(\lambda). \tag{4.311}$$

可推出

$$\begin{aligned}\mathrm{val}(\mathrm{DQ}_h^2) &\geqslant \sup_{\lambda\in\mathcal{S}(\mathrm{DL}_d)} \xi_h^{2,R}(\lambda).\\ &= 2\mathrm{val}(\mathrm{PL}_d) + \sup_{\lambda\in\mathcal{S}(\mathrm{DL}_d)} D_{xx}^2 L(x_0,\lambda,u_0)(h,h).\end{aligned}$$

将此式与 (4.308) 相结合, 得到 $\mathrm{val}(\mathrm{DQ}_h^2)\geqslant 2\mathrm{val}(\mathrm{PL}_d)$. 这一不等式连同命题 4.108 与引理 4.110(v), 证得 (i).

(ii) 由引理 4.110(v) 与 (4.311), 只需证明 $\mathrm{val}(\mathcal{DQ}^2) \leqslant \mathrm{val}(\mathcal{DQ}^{2,R})$. 若条件 (a) 成立, 则对给定的 $\lambda \in \Lambda(x_0, u_0)$, 由 $\sigma(\lambda, \mathcal{T}_K^2(h)) \leqslant 0$, 有 $\sigma(\lambda, \mathcal{T}_K^2(h)) = 0$, 从而 $\xi_h^{2,R}(\lambda) = \xi_h^2(\lambda)$ 对每一临界方向 h 成立. 结果, 问题 $(\mathcal{DQ}^2)$ 与 $(\mathcal{DQ}^{2,R})$ 具有相同的最优值, 这就完成了这种情形的证明.

现在设 (b) 成立. 由定理 3.9, Lagrange 乘子集合是 Y^* 的有界的弱 * 紧致子集. 类似于命题 4.89 的证明, 得到函数 $h \mapsto \sup\limits_{\lambda \in \Lambda(x_0, u_0)} \xi_h^{2,R}(\lambda)$ 是连续的. 令 h_n 是 $(\mathcal{DQ}^{2,R})$ 的极小化序列, 由于广义的多面性条件成立, 对每一 n, 存在临界方向 $\hat{h}_n$, 它与 h_n 充分接近, 满足 $0 \in \mathcal{T}_K^2(\hat{h}_n)$ 且

$$
\begin{aligned}
\mathrm{val}(\mathrm{DQ}^2_{h_n}) &= \sup_{\lambda \in \Lambda(x_0, u_0)} \xi_{h_n}^{2,R}(\lambda) \geqslant \sup_{\lambda \in \Lambda(x_0, u_0)} \xi_{\hat{h}_n}^{2,R}(\lambda) - n^{-1} \\
&= \mathrm{val}(\mathrm{DQ}^2_{\hat{h}_n}) - n^{-1}.
\end{aligned}
$$

因为 h_n 是问题 $(\mathcal{DQ}^{2,R})$ 的极小化序列, 可推出不等式 $\mathrm{val}(\mathcal{DQ}^2) \leqslant \mathrm{val}(\mathcal{DQ}^{2,R})$, 这就完成此种情况的证明.

(iii) 设 x_0 满足二阶充分条件 (3.136). 因为 $\Lambda(x_0, u_0)$ 是非空的, 由引理 3.65, 存在 $\beta > 0$ 与 $M > 0$, 满足对每一临界方向 h, 存在 $\lambda \in \Lambda(x_0, u_0) \cap B(0, M)$, 使得 $D_{xx}^2 L^g(x_0, \alpha, \lambda)(h, h) \geqslant \beta \|h\|^2$, 可得到

$$
\begin{aligned}
\sup_{\lambda \in \Lambda(x_0, u_0)} \xi_h^{2,R}(\lambda) &\geqslant \sup_{\lambda \in \Lambda(x_0, u_0) \cap B(0, M)} \xi_h^{2,R}(\lambda) \\
&\geqslant 2 D_u f(x_0, u_0) d - M \|D_u g(x_0, u_0) d\| + \beta \|h\|^2,
\end{aligned}
$$

因此有

$$
\mathrm{val}(\mathcal{DQ}^{2,R}) \geqslant 2 D_u f(x_0, u_0) d - M \|D_u g(x_0, u_0) d\|
$$

是下方有界的. 证毕 □

由上述命题中的结论 (i) 可得, 在二阶条件 (4.308) 之下, 命题 4.108 与命题 4.22 得到的上方估计是相同的.

4.8.2 最优解的下估计与展式

这一节基于由 (4.307) 定义的松弛问题 $(\mathcal{DQ}^{2,R})$ 推导最优值函数的下方估计.

定理 4.112 设 X 是自反的 Banach 空间. 设

(i) $\Lambda(x_0, u_0)$ 是非空的且方向正则性条件成立.

(ii) $(\mathrm{P}_{u(t)})$ 具有 $o(t)$ 最优解 $\bar{x}(t)$, 且满足 $\|\bar{x}(t) - x_0\| = O(t^{1/2})$, $t \geqslant 0$.

(iii) 对所有的 $\lambda \in \Lambda(x_0, u_0)$, 函数 $D_{xx}^2 L(x_0, \lambda, u_0)(\cdot, \cdot)$ 弱下半连续, 则

$$
\liminf_{t \downarrow 0} \frac{v(u(t)) - v(u_0)}{t} \geqslant \frac{1}{2} \mathrm{val}(\mathcal{DQ}^{2,R}). \tag{4.312}
$$

证明 类似于定理 4.91 的证明. 因为 X 是自反的, 比率 $h(t) := t^{-1/2}(\bar{x}(t)-x_0)$ 是有界的, 它至少有一弱极限点 $\bar{h}$. 由

$$T_K(y_0) \ni \frac{G(\bar{x}(t),u(t))-G(x_0,u_0)}{t^{1/2}} = D_xG(x_0,u_0)h(t)+o(1)$$

及 $T_K(y_0)$ 的弱闭性, 有 $\bar{h} \in T_K(y_0)$. 也由于 $\bar{x}(t)$ 是 $o(t)$ 最优的,

$$v(u(t)) = f(\bar{x}(t),u(t))+o(t) = v(u_0)+t^{1/2}D_xf(x_0,u_0)h(t)+o(t^{1/2}).$$

由命题 4.22 可得 $v(u(t)) \leqslant v(u_0)+O(t)$, 有 $D_xf(x_0,u_0)\bar{h} \leqslant 0$, 因此 $\bar{h}$ 是临界方向, 则由 (4.229), 对任何 $\lambda \in \Lambda(x_0,u_0)$ 有

$$\begin{aligned} v(u(t)) &= f(\bar{x}(t),u(t))+o(t) \\ &\geqslant f(x_0,u_0)+L(x(t),\lambda,u(t))-L(x_0,\lambda,u_0)+o(t) \\ &= v(u_0)+\frac{1}{2}t\xi_{h(t)}^{2,R}(\lambda)+o(t). \end{aligned}$$

设存在序列 $t_n \downarrow 0$, $\bar{h}$ 是 $h(t_n)$ 的弱极限点, 满足 $t^{-1}(v(u(t))-v(u_0))$ 沿这一序列取到其下极限. 由假设 (iii) 有

$$\xi_{\bar{h}}^{2,R}(\lambda) \leqslant \liminf_{n\to\infty}\xi_{h(t_n)}^{2,R}(\lambda) \leqslant \liminf_{t\downarrow 0}\frac{v(u(t))-v(u_0)}{\frac{1}{2}t}. \tag{4.313}$$

因为 $\bar{h} \in C(x_0)$, 对 $\lambda \in \Lambda(x_0,u_0)$ 取极大化得到

$$\liminf_{t\downarrow 0}\frac{v(u(t))-v(u_0)}{\frac{1}{2}t} \geqslant \mathrm{val}(\mathrm{DQ}_{\bar{h}}^{2,R}) \geqslant \mathrm{val}(\mathcal{D}Q^{2,R}). \tag{4.314}$$

结论得证. □

上述结果使我们在假设广义多面性条件成立的前提下, 得到最优值函数的方向导数, 在某些更强的假设之下, 得到最优解的方向展开.

定理 4.113 令 X 是自反的 Banach 空间. 设

(i) $\Lambda(x_0,u_0)$ 非空且方向正则性条件成立.

(ii) 二阶充分性条件 (3.136) 在 $x_0 \in \Phi(u_0)$ 处成立.

(iii) $(\mathrm{P}_{u(t)})$ 具有 $o(t)$ 最优解 $\bar{x}(t)$, 且满足当 $t \downarrow 0$ 时 $\bar{x}(t) \to x_0$.

(iv) 对所有 $\lambda \in \Lambda(x_0,u_0)$, 函数 $D_{xx}^2L(x_0,\lambda,u_0)(\cdot,\cdot)$ 是弱下半连续的.

(v) 或者 (a) 对每一临界方向 $h \in C(x_0)$, $0 \in \mathcal{T}_K^2(h)$, 或者 (b) Robinson 约束规范 (2.163) 与广义多面性条件成立.

则

$$\|\bar{x}(t)-x_0\| = O(t^{1/2}), \tag{4.315}$$

$h(t) := t^{-1/2}(\bar{x}(t)-x_0)$ 的任何弱极限点 $\bar{h}$ 是 $(\mathcal{D}Q^{2,R})$ 的最优解, 最优值函数的下述展式成立:

$$v(u(t)) = v(u_0) + \frac{1}{2}t\ \mathrm{val}(\mathcal{D}Q^{2,R}) + o(t). \tag{4.316}$$

设还有下述两个条件成立:

(vi) 对所有的 $\lambda \in \Lambda(x_0,u_0)$, 函数 $D^2_{xx}L(x_0,\lambda,u_0)(\cdot,\cdot)$ 是 Legendre 形式.

(vii) 存在 $\overline{\lambda} \in \Lambda(x_0,u_0)$ 满足对某一 $h(t)$ 的弱极限点 $\bar{h}$ 有 $\xi^{2,R}_{\bar{h}}(\overline{\lambda}) = \mathrm{val}(\mathrm{DQ}^2_{\bar{h}})$. 则 $\bar{h}$ 是 $h(t)$ 的强极限点, 且是 $(\mathcal{D}Q^{2,R})$ 的最优解. 尤其, 若 $(\mathcal{D}Q^{2,R})$ 有唯一的最优解 $\hat{h}$, 则 $\hat{h}=\bar{h}$ 且对 $(\mathrm{P}_{u(t)})$ 的任何 $o(t)$ 最优路径 $\hat{x}(t)\to x_0$, 下述展式成立①

$$\hat{x}(t) = x_0 + t^{\frac{1}{2}}\hat{h} + o(t^{\frac{1}{2}}). \tag{4.317}$$

证明 由定理 4.53, Hölder 稳定性 (4.315) 是假设 (i) 与 (ii) 的一个结果. 结合命题 4.108, 命题 4.111(ii) 与定理 4.112, 得到 (4.316). 重复定理 4.112 的证明, 结合 (4.316), 由 (4.314) 得到 $\bar{h}$ 是 $(\mathcal{D}Q^{2,R})$ 的最优解, 由 (vii) 与 (4.313) 得

$$\xi^{2,R}_{\bar{h}}(\overline{\lambda}) = \mathrm{val}(DQ^2_{\bar{h}}) = \limsup_{n\to\infty}\xi^{2,R}_{h(t_n)}(\overline{\lambda}).$$

由 (vi) 得到 $\bar{h}$ 是 $h(t)$ 的强极限点. 若 $(\mathcal{D}Q^{2,R})$ 有唯一最优解 $\hat{h}$, 则 $\hat{h}=\bar{h}$ 是 $h(t)$ 的极限, 由此可得 (4.317). □

注 4.114 若 Robinson 约束规范 (2.163) 成立, 则上述定理中的假设 (iii) 成立, 因为由定理 2.6, 在弱 * 紧致集 $\Lambda(x_0,u_0)$ 上极大化弱 * 连续的目标函数 $\xi^{2,R}_{\bar{h}}(\cdot)$ 的问题有一最优解.

若问题 (P) 是有限约束的, 假设 (vii) 也是成立的, 因为有限值线性规划问题有一最优解 (见定理 2.202).

4.8.3 Lagrange 乘子空集

设 x_0 是非扰动问题 (P) 的局部最优解, (P) 按 (4.5) 的形式给出, 令 $d\in U$, 使得方向正则性条件 (见定义 4.7) 成立. 令人惊奇的是, 即使 Lagrange 乘子不存在时也可以进行二阶分析. 此种情形时的分析基于奇异 Lagrange 乘子. 回顾在第 3 章中关于广义 Lagrange 乘子讨论的某些结果.

命题 3.16 已经证明, (对问题 (P) 在点 x_0 处) 奇异 Lagrange 乘子存在的充要条件是集合

$$\mathcal{Z}(x_0) := DG(x_0)X - \mathcal{R}_K(G(x_0))$$

① 原著中为 $\hat{x}$.

的闭包不等于 Y. 因为 Robinson 约束规范等价于 $\mathcal{Z}(x_0)=Y$, 广义 Lagrange 乘子集合是空集的必要条件是凸锥 $\mathcal{Z}(x_0)$ 不同于 Y, $\mathcal{Z}(x_0)$ 是 Y 的稠密子集合. 这种情形只有在 $\mathcal{Z}(x_0)$ 具有空的相对内部时发生, 见命题 3.16. 若集合 K 具有非空的内部, 则集合 $\mathcal{Z}(x_0)$ 具有非空的内部, 因此, 此种情形下存在广义 Lagrange 乘子. 进一步, 此种情形下由定理 3.50 给出的二阶必要条件 (3.116) 是成立的. 这些必要条件是基于广义 Lagrange 乘子的, 不要求约束规范成立.

最后, 若二阶充分性条件 (3.136) 成立, 则由定理 4.53, 即使在 Lagrange 乘子不存在的情况下, 也可得到 $O(t)$ 最优路径的 $\left(\frac{1}{2}\text{ 度的 }\right)$ 方向 Hölder 稳定性. 在这些情况下, 相应的最优值函数的变化至多是 $O(t^{\frac{1}{2}})$ 阶的. 在 4.8.2 节中已经看到, 在合理的假设下, 只要 Lagrange 乘子存在, 最优值函数的变化是 $O(t)$ 阶的. 相对地, 下述引理表明, 在 Lagrange 乘子不存在的情况下, 最优值函数以 t 的平方根的变化率典型地变化.

引理 4.115 *令 $x_0 \in \mathcal{S}(u_0)$. 设在 x_0 处沿方向 d 的方向正则性条件成立, 存在 $\bar{h} \in C(x_0)$ 满足*

$$D_x f(x_0,u_0)\bar{h} < 0 \text{ 且 } \mathcal{T}_K^2(\bar{h}) \neq \varnothing, \tag{4.318}$$

则

$$\limsup_{t\downarrow 0} \frac{v(u(t))-v(u_0)}{t^{\frac{1}{2}}} < 0. \tag{4.319}$$

证明 令 $\alpha>0, h:=\alpha\bar{h}$. 因为 $\mathcal{T}_K^2(\bar{h})$ 是非空的, 存在 $k\in\mathcal{T}_K^2(\bar{h})$, 因此有 $\alpha^2 k\in\mathcal{T}_K^2(h)$. 考虑集合

$$E:=\mathcal{T}_K^2(h)-DG(x_0,u_0)(X,2d)-D_{xx}^2G(x_0,u_0)(h,h).$$

由 (4.13), 存在 $\varepsilon>0$ 满足 $2\varepsilon B_Y\subset T_K(y_0)-DG(x_0,u_0)(X,2d)$. 与 (3.62) 相结合得到, 对 $\alpha>0$ 充分小,

$$\varepsilon B_Y\subset T_K(y_0)-DG(x_0,u_0)(X,2d)+\alpha^2[k-D_{xx}^2G(x_0,u_0)(h,h)]\subset E.$$

集合 E 包含 0 可推出存在 $w\in X$ 满足

$$2DG(x_0,u_0)(w,d)+D_{xx}^2G(x_0,u_0)(h,h)\in\mathcal{T}_K^2(h). \tag{4.320}$$

考虑路径

$$x(t):=x_0+t^{\frac{1}{2}}h+tw. \tag{4.321}$$

由 $G(\cdot,\cdot)$ 在 (x_0,u_0) 处的二阶 Taylor 展式 (4.288) 和二阶切集的定义有

$$\operatorname{dist}(G(x(t),u(t)),K)=o(t).$$

则由引理 4.109, 对充分小的 $\theta > 0$, 存在路径 $\bar{x}(\cdot) : \mathbb{R}_+ \to X$ 满足

$$G(\bar{x}(t), u((1+\theta)t)) \in K \text{ 且 } \|\bar{x}(t) - x(t)\| \leqslant c\theta t.$$

于是有

$$\begin{aligned} v(u((1+\theta)t)) &\leqslant f(\bar{x}(t), u((1+\theta)t)) \\ &\leqslant f(x_0, u_0) + t^{\frac{1}{2}} D_x f(x_0, u_0)h + o(t^{\frac{1}{2}}), \end{aligned}$$

因而得到

$$\limsup_{t \downarrow 0} \frac{v(u(t)) - v(u_0)}{t^{\frac{1}{2}}} \leqslant (1+\theta)^{-\frac{1}{2}} D_x f(x_0, u_0)h < 0, \tag{4.322}$$

证毕. □

注 4.116 由命题 3.44, 若 Lagrange 乘子集合 $\Lambda(x_0, u_0)$ 非空, 则对所有的临界方向 $h \in C(x_0)$, 有 $Df(x_0, u_0)h = 0$. 所以, 由满足 $D_x f(x_0, u_0)\bar{h} < 0$ 的临界方向 $\bar{h}$ 的存在性可推出集合 $\Lambda(x_0, u_0)$ 是空集. 注意到, 不等式 (4.319) 可以表示为下述形式

$$v(u(t)) - v(u_0) \leqslant -\varepsilon t^{\frac{1}{2}},$$

其中 $\varepsilon > 0$ 是常数, 对所有充分小的 $t \geqslant 0$ 成立. 上述不等式表明 $v(u(t))$ 至少以 $\sqrt{t}$ 的速率变化.

因为对 (4.321) 形式的路径, 目标函数展式的第一项是 $D_x f(x_0, u_0)h$, 因为由引理 4.115, 这一项通常而言是非零的, 很自然考虑受约束于存在 w 满足 $G(x(t), u(t))$ 到 K 的距离是 $o(t)$ 阶的极小化这一项的问题, 即约束为

$$2DG(x_0, u_0)(w, d) + D_{xx}^2 G(x_0, u_0)(h, h) \in \mathcal{T}_K^2(h).$$

考虑下述优化问题

$$(\mathcal{P}Q^3) \qquad \begin{aligned} &\min_{h \in C(x_0)} && D_x f(x_0, u_0)h \\ &\text{s. t.} && D_{xx}^2 G(x_0, u_0)(h, h) \in \mathcal{T}_K^2(h) - 2DG(x_0, u_0)(X, d). \end{aligned} \tag{4.323}$$

注意到上述极小化问题仅用变量 h 表示, 不涉及 w.

命题 4.117 令 $x_0 \in \mathcal{S}(u_0)$, 设在 x_0 处沿方向 d 处的方向正则性条件成立, 则

$$\limsup_{t \downarrow 0} \frac{v(u(t)) - v(u_0)}{t^{\frac{1}{2}}} \leqslant \operatorname{val}(\mathcal{P}Q^3). \tag{4.324}$$

证明 令 h 是问题 $(\mathcal{P}Q^3)$ 的可行点, 则存在 w 满足 (4.320). 应用引理 4.115 的推证于由 (4.321) 定义的 $x(t)$ 得到, 对 $\theta > 0$ 充分小, 有 (4.322) 成立. 令 $\theta \downarrow 0$, 因为 h 是 $(\mathcal{P}Q^3)$ 的任意可行点, 结论得证. □

为了理解问题 $(\mathcal{P}\mathcal{Q}^3)$ 的对偶形式, 对固定的 $h\in C(x_0)$, 下面叙述相应的问题. 即

$$(PQ_h^3)\qquad \begin{aligned} &\min_{w\in X} && D_xf(x_0,u_0)h,\\ &\text{s. t.} && 2DG(x_0,u_0)(w,d)+D_{xx}^2G(x_0,u_0)(h,h)\in\mathcal{T}_K^2(h). \end{aligned}\tag{4.325}$$

若上述问题是相容的, 则其最优值是 $D_xf(x_0,u_0)h$, 否则, 其最优值是 $+\infty$. 正式地, (PQ_h^3) 的对偶可以表示为下述形式

$$(DQ_h^3)\qquad \begin{aligned} D_xf(x_0,u_0)h+\max_{\lambda\in\Lambda^s(x_0,u_0)}&\{2D_uL^s(x_0,\lambda,u_0)d\\ &+D_{xx}^2L^s(x_0,\lambda,u_0)(h,h)-\sigma(\lambda,\mathcal{T}_K^2(h))\}, \end{aligned}\tag{4.326}$$

其中 $\Lambda^s(x_0,u_0)$ 记奇异 Lagrange 乘子集合, $L^s(x,\lambda,u)$ 记奇异 Lagrange 函数

$$L^s(x,\lambda,u):=\langle\lambda,G(x,u)\rangle.$$

设 $\Lambda^s(x_0,u_0)\neq\varnothing$, 若对所有的 $\lambda\in\Lambda^s(x_0,u_0)$,

$$2D_uL^s(x_0,\lambda,u_0)d+D_{xx}^2L^s(x_0,\lambda,u_0)(h,h)\leqslant\sigma(\lambda,\mathcal{T}_K^2(h)),$$

则上述问题的最优值是 $D_xf(x_0,u_0)h$, 否则, 最优值是 $+\infty$. 因此, $(\mathcal{P}\mathcal{Q}^3)$ 的对偶问题为

$$(\mathcal{D}\mathcal{Q}^30)\qquad \begin{aligned} &\min_{h\in C(x_0)} && D_xf(x_0,u_0)h,\\ &\text{s. t.} && 2D_uL^s(x_0,\lambda,u_0)d+D_{xx}^2L^s(x_0,\lambda,u_0)(h,h)\leqslant\sigma(\lambda,\mathcal{T}_K^2(h)),\\ &&& \text{对所有的 }\lambda\in\Lambda^s(x_0,u_0). \end{aligned}$$

这一问题的等价形式为

$$(\mathcal{D}\mathcal{Q}_*^3)\qquad \begin{aligned} &\min_{h\in C(x_0)} && D_xf(x_0,u_0)h,\\ &\text{s. t.} && D_{xx}^2G(x_0,u_0)(h,h)\in\mathrm{cl}\{\mathcal{T}_K^2(h)-2DG(x_0,u_0)(X,d)\}. \end{aligned}$$

这一等价性可由下述引理刻画.

引理 4.118　设奇异 Lagrange 乘子集 $\Lambda^s(x_0,u_0)$ 是非空的, 则下述结论成立:

(i) 下述条件是等价的:

$$y\in\mathrm{cl}\{\mathcal{T}_K^2(h)-2D_xG(x_0,u_0)X\},\tag{4.327}$$

$$\langle\lambda,y\rangle\leqslant\sigma(\lambda,\mathcal{T}_K^2(h)),\quad \forall\lambda\in\Lambda^s(x_0,u_0).\tag{4.328}$$

(ii) 问题 $(\mathcal{P}\mathcal{Q}^3)$, $(\mathcal{D}\mathcal{Q}^3)$ 与 $(\mathcal{D}\mathcal{Q}^3_*)$ 具有相同的最优值.

证明 若 $\mathcal{T}^2_K(h)$ 是空集, 则 (4.327) 与 (4.328) 两个条件都不正确. 因此, 设 $\mathcal{T}^2_K(h) \neq \varnothing$. 显然, 由 (4.327) 可推出 (4.328). 为证明相反的结果, 假设 (4.327) 不成立, 则由分离定理(定理 2.14), 存在 $\lambda \in Y^* \setminus \{0\}$ 与 $\alpha \in \mathbb{R}$ 满足

$$\langle \lambda, y \rangle > \alpha \geqslant \langle \lambda, k - D_x G(x_0, u_0) w \rangle$$

对所有 $k \in \mathcal{T}^2_K(h)$ 与 $w \in X$ 成立. 对 $w \in X$ 取极大, 得到 $D_x G(x_0, u_0)^* \lambda = 0$, 再对 k 取上确界, 得到

$$\langle \lambda, y \rangle > \alpha \geqslant \sigma(\lambda, \mathcal{T}^2_K(h)). \tag{4.329}$$

因为由 (3.62) 可得到 $\mathcal{T}^2_K(h) \neq \varnothing$, 它的回收锥包含 $T_K(G(x_0, u_0))$, 因此 (3.62) 可推出 $\lambda \in N_K(G(x_0, u_0))$. 所以, $\lambda \in \Lambda^s(x_0, u_0)$, (4.329) 与 (4.328) 相矛盾. 这完成了结论 (i) 的证明.

(ii) 由 (i) 得

$$\operatorname{val}(DQ^3_h) = \begin{cases} D_x f(x_0, u_0) h, & \text{若 } 2D_u G(x_0, u_0) d + D^2_{xx} G(x_0, u_0)(h, h) \\ & \quad \in \operatorname{cl}\{\mathcal{T}^2_K(h) - 2D_x G(x_0, u_0) X\}, \\ +\infty, & \text{否则}. \end{cases}$$

在任何情况下, 都有 $\operatorname{val}(DQ^3_h) \geqslant D_x f(x_0, u_0) h$, 因此, 若原始问题是相容的, 则原始与对偶最优值相同. 令 $\bar{h} \in C(x_0)$ 对 $(\mathcal{D}\mathcal{Q}^3)$ 是可行. 给定 $\eta < 1$, 置 $h := \eta \bar{h}$. 则由与引理 4.110(ii) 相同的推证得到, 问题 (4.325) 与 (4.326) 具有相同的有限的值 $D_x f(x_0, u_0) h = \eta D_x f(x_0, u_0) \bar{h}$, 因此

$$\operatorname{val}(\mathcal{P}\mathcal{Q}^3) \leqslant \limsup_{\eta \uparrow 1} [\operatorname{val}(DQ^3_{\eta \bar{h}})] = D_x f(x_0, u_0) \bar{h} = \operatorname{val}(DQ^3_{\bar{h}}).$$

于是 $\operatorname{val}(\mathcal{P}\mathcal{Q}^3) \leqslant \operatorname{val}(\mathcal{D}\mathcal{Q}^3)$. 相反的不等式由 $\operatorname{val}(DQ^3_h) \leqslant \operatorname{val}(PQ^3_h)$ 得到, 结论 (ii) 得证. □

像当广义多面性条件成立时推导最优解的展式那样, 现在讨论去掉 sigma 项时得到的下方估计. 像通常一样, 置 $y_0 := G(x_0, u_0)$. 如上所述, 对松弛问题, 得到两个等价的表示

$$(\mathcal{D}\mathcal{Q}^{3,R}) \qquad \begin{array}{ll} \min\limits_{h \in C(x_0)} & D_x f(x_0, u_0) h, \\ \text{s. t.} & 2D_u L^s(x_0, \lambda, u_0) d + D^2_{xx} L^s(x_0, \lambda, u_0)(h, h) \leqslant 0, \\ & \text{对所有的 } \lambda \in \Lambda^s(x_0, u_0) \end{array}$$

与

$$(\mathcal{D}\mathcal{Q}^{3,R}_*) \qquad \begin{array}{ll} \min\limits_{h \in C(x_0)} & D_x f(x_0, u_0) h, \\ \text{s. t.} & D^2_{xx} G(x_0, u_0)(h, h) \in \operatorname{cl}\{T_K(y_0) - 2DG(x_0, u_0)(X, d)\}. \end{array}$$

由于

$$T_{T_K(y_0)}(D_xG(x_0,u_0)h)=\mathrm{cl}\{T_K(y_0)+D_xG(x_0,u_0)h\},$$

有

$$\mathrm{cl}\{T_KG(x_0,u_0)-2DG(x_0,u_0)(X,d)\}=\mathrm{cl}\{T_{T_K(y_0)}(D_xG(x_0,u_0)h)-2DG(x_0,u_0)(X,d)\}.$$

当 sigma 项消失时, 作为引理 4.118 的结论, 问题 $(\mathcal{D}Q^{3,R})$ 与 $(\mathcal{D}Q_*^{3,R})$ 是等价的.

定理 4.119 令 X 是自反的 Banach 空间, 设

(i) Lagrange 乘子集合 $\Lambda(x_0,u_0)$ 是空集, $\Lambda^s(x_0,u_0)\neq\varnothing$, 方向正则性条件成立.

(ii) $(\mathrm{P}_{u(t)})$ 具有 $o(t)$ 最优解 $\bar{x}(t)$ 满足 $\|\bar{x}(t)-x_0\|=O(t^{\frac{1}{2}})$, $t\geqslant 0$.

(iii) 对所有的 $\lambda\in\Lambda(x_0,u_0)$, 函数 $D_{xx}^2L(x_0,\lambda,u_0)(\cdot,\cdot)$ 是弱下半连续的.

则

$$\liminf_{t\downarrow 0}\frac{v(u(t))-v(u_0)}{t^{\frac{1}{2}}}\geqslant \mathrm{val}(\mathcal{D}Q^{3,R}). \tag{4.330}$$

证明 因为 $h(t):=t^{-\frac{1}{2}}(\bar{x}(t)-x_0)$ 是有界的, 存在序列 $t_n\downarrow 0$ 满足 $t^{-\frac{1}{2}}(v(u(t))-v(u_0))$ 取得其下极限, 有极限点 $\bar{h}$. 显然, 这一下极限等于 $Df(x_0,u_0)\bar{h}$. 另一方面, 由于 $\Lambda^s(x_0,u_0)\subset N_K(G(x_0,u_0))$, 对每一 $\lambda\in\Lambda^s(x_0,u_0)$, 有

$$\begin{aligned}0&\geqslant L^s(x(t_n),\lambda,u(t_n))-L^s(x_0,\lambda,u_0)\\&=\frac{1}{2}t_n(2D_uL(x_0,\lambda,u_0)d+D_{xx}^2L(x_0,\lambda,u_0)(\bar{h},\bar{h}))+o(t_n)\quad(\text{原著中为 }o(t)).\end{aligned}$$

因此

$$2D_uL(x_0,\lambda,u_0)d+D_{xx}^2L(x_0,\lambda,u_0)(\bar{h},\bar{h})\leqslant 0.$$

由于这一不等式对每一 $\lambda\in\Lambda^s(x_0,u_0)$ 均成立, 因而有 $Df(x_0,u_0)\bar{h}\geqslant\mathrm{val}(\mathcal{D}Q^{3,R})$. 结论得证. □

本节的最后一个定理讨论最优值函数的 $O(t^{\frac{1}{2}})$ 阶的展开可以计算的情形.

定理 4.120 设 X 是自反的 Banach 空间. 设

(i) Lagrange 乘子集合 $\Lambda(x_0,u_0)$ 是空集, 而 $\Lambda^s(x_0,u_0)\neq\varnothing$, 且方向正则性条件成立.

(ii) 二阶充分性条件 (3.136) 在 $x_0\in\Phi(u_0)$ 处成立.

(iii) $(\mathrm{P}_{u(t)})$ 具有 $o(t)$ 最优解 $\bar{x}(t)$, 且满足当 $t\downarrow 0$ 时 $\bar{x}(t)\to x_0$.

(iv) 对所有的 $\lambda\in\Lambda^s(x_0,u_0)$, 函数 $D_{xx}^2L^s(x_0,\lambda,u_0)(\cdot,\cdot)$ 是弱下半连续的.

(v) 对所有的临界方向 $h\in C(x_0)$, $0\in T_K^2(h)$.

则

$$\|\bar{x}(t) - x_0\| = O(t^{\frac{1}{2}}),$$

$h(t) := t^{-\frac{1}{2}}(\bar{x}(t) - x_0)$ 的任何弱极限 $\bar{h}$ 均是 $(\mathcal{D}Q^{3,R})$ 的最优解, 最优值函数的下述展式成立:

$$v(u(t)) = v(u_0) + t^{\frac{1}{2}}\mathrm{val}(\mathcal{D}Q^{3,R}) + o(t^{\frac{1}{2}}). \tag{4.331}$$

尤其, 若 $(\mathcal{D}Q^{3,R})$ 具有唯一的最优解 $\hat{h}$, 且 X 是有限维的, 则 $\hat{h} = \bar{h}$, 且对 $(\mathrm{P}_{u(t)})$ 的任何 $o(t)$ 最优路径, 下述展开式成立①:

$$\hat{x}(t) = x_0 + t^{\frac{1}{2}}\hat{h} + o(t^{\frac{1}{2}}). \tag{4.332}$$

证明 结合命题 4.117 与定理 4.119, 由 (v) 可得问题 $(\mathcal{D}Q^3)$ 与 $(\mathcal{D}Q^{3,R})$ 具有相同的最优值, 从而可得展式 (4.331). 在 X 是有限维空间的情形, 则 $(\mathcal{D}Q^{3,R})$ 具有唯一解, 由紧致性推证可得 (4.332). □

4.8.4 二阶正则问题的 Hölder 展开式

这一节讨论当 X 是有限维空间时强正则性(见定义 3.85) 成立的情形. 此时可以得到目标函数和 (近似) 最优解的展式.

定理 4.121 设

(i) X 是有限维空间.

(ii) $(\mathrm{P}_{u(t)})$ 的最优解集非空, 且对 $t > 0$ 充分小是一致有界的.

(iii) 非扰动问题 (P_{u_0}) 具有唯一最优解 x_0, 满足 $\Lambda(x_0, u_0) \neq \varnothing$.

(iv) 方向正则性条件在 x_0 处沿方向 d 成立.

(v) $\mathcal{A}(h) := \mathcal{T}_K^2(h)$ 的二阶充分性条件 (3.160) 成立.

(vi) 对每一 $h \in C(x_0)$, 集合 K 在 $G(x_0, u_0)$ 处沿方向 $D_xG(x_0, u_0)h$ 关于 $D_xG(x_0, u_0)$ 是二阶正则的.

则 $(\mathrm{P}_{u(t)})$ 的任何 $o(t)$ 最优解路径 $\bar{x}(t)$ 均是 $\frac{1}{2}$ 度 Hölder 稳定的, $t^{-\frac{1}{2}}(\bar{x}(t) - x_0)$ 的每一极限点均是 $(\mathcal{D}Q^2)$ 的最优解, 且最优值函数可展开为

$$v(u(t)) = v(u_0) + \frac{1}{2}t\,\mathrm{val}(\mathcal{D}Q^2) + o(t). \tag{4.333}$$

若还有, $(\mathcal{D}Q^2)$ 具有最优解 $\bar{h}$ 且 Robinson 约束规范成立, 则存在 $(\mathrm{P}_{u(t)})$ 的 $o(t)$ 最优路径 $x(t)$ 满足 $x(t) = x_0 + t^{\frac{1}{2}}\bar{h} + o(t^{\frac{1}{2}})$.

证明 $\bar{x}(t)$ 是 $\frac{1}{2}$ 度 Hölder 稳定的可由定理 4.53 得到. 由命题 4.108, (4.333) 的左端小于或等于其右端. 下证相反的不等式. 令 $t_k \downarrow 0$, $x_k := \bar{x}(t_k)$. 可设 $t_k^{-\frac{1}{2}}(x_k - x_0)$

① 原著中为 $\hat{x}$.

收敛到某 $h \in X$, 因而 $x_k = x_0 + t_k^{\frac{1}{2}} h + t_k z_k$, 满足 $t_k^{\frac{1}{2}} z_k \to 0$. 则 $D_x G(x_0, u_0) h \in T_K(G(x_0, u_0))$, 由 K 的二阶正则性, 有

$$2DG(x_0, u_0)(z_k, d) + D_{xx}^2 G(x_0, u_0)(h, h) + o(1) \in \mathcal{T}_K^2(h). \tag{4.334}$$

由

$$\begin{aligned} f(x_k, u(t_k)) &= v(u(t_k)) + o(t_k) \leqslant v(u_0) + t_k \mathrm{val}(\mathrm{PLd}) + o(t_k) \\ &= f(x_0, u_0) + o(t_k^{\frac{1}{2}}), \end{aligned}$$

有 $D_x f(x_0, u_0) h \leqslant 0$, 因此 $h \in C(x_0)$.

令 $\lambda \in \Lambda(x_0, u_0)$[①]. 由 $D_x L(x_0, \lambda, u_0) = 0$, $\langle \lambda, D_x G(x_0, u_0) h \rangle = 0$ 与二阶正则性得到

$$\begin{aligned} v(u(t_k)) &= f(x_k, u(t_k)) + o(t_k) \\ &= f(x_0, u_0) + L(x_k, \lambda, u(t_k)) - L(x_0, \lambda, u_0) \\ &\quad - \langle \lambda, G(x_k, u(t_k)) - G(x_0, u_0) \rangle + o(t_k) \\ &= v(u_0) + \frac{1}{2} t_k [2D_u L(x_0, \lambda, u_0) d + D_{xx}^2 L(x_0, \lambda, u_0)(h, h)] \\ &\quad - \frac{1}{2} t_k \langle \lambda, 2DG(x_0, u_0)(z_k, d) + D_{xx}^2 L(x_0, \lambda, u_0)(h, h) \rangle + o(t_k) \\ &\geqslant v(u_0) + \frac{1}{2} t_k [2D_u L(x_0, \lambda, u_0) d + D_{xx}^2 L(x_0, \lambda, u_0)(h, h) \\ &\quad - \sigma(\lambda, \mathcal{T}_K^2(h))] + o(t_k). \end{aligned}$$

由于上式对任何 $\lambda \in \Lambda(x_0, u_0)$[②]均是成立的, 得到

$$\begin{aligned} &\liminf_{t \downarrow 0} \frac{v(u(t_k)) - v(u_0)}{\frac{1}{2} t_k} \\ &\geqslant \sup_{\lambda \in \Lambda(x_0, u_0)} \{2D_u L(x_0, \lambda, u_0) d + D_{xx}^2 L(x_0, \lambda, u_0)(h, h) - \sigma(\lambda, \mathcal{T}_K^2(h))\}. \end{aligned}$$

与命题 4.108 的上方估计相结合, 可得到 $h \in \mathcal{S}(\mathcal{D}\mathcal{Q}^2)$. 因为 $\Lambda(x_0, u_0) \neq \varnothing$, 且 $\sigma(\lambda, \mathcal{T}_K^2(h)) \leqslant 0$, 所以 $\mathrm{val}(\mathcal{D}\mathcal{Q}^2)$ 是有限的, (4.333) 得证.

剩下要证结论的最后一部分. 由 Robinson 约束规范, 存在形式为 $x(t) = x_0 + t^{\frac{1}{2}} \bar{h} + t\bar{z} + o(t)$ 的可行路径. 计算 $f(x(t), td)$ 的展式, 得 $x(t)$ 是 $o(t)$ 最优路径, 证毕 □

对于 Lagrange 乘子集合为空集的情况, 可给出类似的结论. 因为其证明类似于前述命题的证明, 故从略.

---

① 原著中是令 $\lambda \in \Lambda_0(x_0, u_0)$.

② 原著中为 $\Lambda_0(x_0)$.

定理 4.122 设

(i) 空间 X 是有限维的.

(ii) $(\mathrm{P}_{u(t)})$ 的最优解集非空, 对充分小的 $t>0$ 一致有界.

(iii) 问题 (P_{u_0}) 有唯一最优解 x_0, 满足 $\Lambda(x_0,u_0)=\varnothing$, 但 $\Lambda^s(x_0,u_0)\neq\varnothing$.

(iv) 方向正则性条件成立.

(v) $\mathcal{A}(h):=\mathcal{T}_K^2(h)$ 的二阶充分性条件 (3.160) 成立.

(vi) 对每一 $h\in C(x_0)$, 集合 K 在 $G(x_0,u_0)$ 处沿方向 $D_xG(x_0,u_0)h$ 相对于 $D_xG(x_0,u_0)$ 是二阶正则的.

则 $\bar{x}(t)$ 均是 $\dfrac{1}{2}$ 度 Hölder 稳定的, $t^{-\frac{1}{2}}(\bar{x}(t)-x_0)$ 的任何极限均是问题 $(\mathcal{D}Q^3)$ 的最优解, 且最优值函数可展开为

$$v(u(t))=v(u_0)+t^{\frac{1}{2}}\mathrm{val}(\mathcal{D}Q^3)+o(t^{\frac{1}{2}}). \tag{4.335}$$

4.9 辅助结果

4.9.1 等式约束问题

这一节证明等式约束问题的特殊情况, 稳定性分析可以通过将隐函数定理应用于一阶最优性条件系统而导出. 这与涉及不等式, 或更一般地, 与需要用前面章节中发展的工具来处理的约束问题是不同的. 设 X,Y 与 U 是 Banach 空间. 一个线性连续算子 $H:X\to X^*$ 是自共轭的, 若对所有的 $x,x'\in X$, $\langle Hx,x'\rangle=\langle Hx',x\rangle$; H 称为是非负的, 若对所有的 $x\in X$, $\langle Hx,x\rangle\geqslant 0$; H 称为是可逆的, 若它是一对一的映上的, 且逆算子 H^{-1} 是连续的; 称 H 是一致正的, 若存在 $\alpha>0$ 满足, 对所有的 $x\in X$, $\langle Hx,x\rangle\geqslant\alpha\|x\|^2$. 注意到, 若存在一自共轭的一致正的算子 $H:X\to X^*$, 则 $\|x\|_1:=\langle Hx,x\rangle^{\frac{1}{2}}$ 定义了 X 的范数, 这一范数与原范数是等价的, 因此空间 X 是可 Hilbert 化的, 所以也是自反的.

引理 4.123 令 X 是 Banach 空间, $H\in\mathcal{L}(X,X^*)$ 是自共轭的非负的线性连续算子. 则 H 是可逆的当且仅当它是一致正的.

证明 设 H 是自共轭的非负可逆的. 我们证 H 是一致正的. 给定 $\bar{x}\in X$, 置 $x^*:=H\bar{x}$. 因为 H 是非负的, (二次) 最优化问题

$$\min_{x\in X}\ \frac{1}{2}\langle Hx,x\rangle-\langle x^*,x\rangle \tag{4.336}$$

是凸的. 所以, 最优解可由最优性条件 $Hx=x^*$ 来刻画. 因为 H 是可逆的, 则唯一解为 $\bar{x}=H^{-1}x^*$. 所以

$$-\frac{1}{2}\langle x^*,H^{-1}x^*\rangle\leqslant\frac{1}{2}\langle Hx,x\rangle-\langle x^*,x\rangle,\quad \forall x\in X, \tag{4.337}$$

或等价地

$$\langle H\bar{x},\bar{x}\rangle \geqslant 2\langle x^*,x\rangle - \langle Hx,x\rangle, \quad \forall x\in X. \tag{4.338}$$

存在 $x_1 \in X$ 满足 $\|x_1\|=1$, 且 $\langle x^*,x_1\rangle \geqslant \frac{1}{2}\|x^*\|$. 在 (4.338) 中置 $x:=\gamma x_1$, $\gamma\in\mathbb{R}$, 得

$$\langle H\bar{x},\bar{x}\rangle \geqslant 2\gamma\langle x^*,x_1\rangle - \gamma^2\langle Hx_1,x_1\rangle, \quad \forall \gamma\in\mathbb{R}.$$

在上述不等式的右端取 $\gamma\in\mathbb{R}$ 的极大化, 由 $\|\bar{x}\| = \|H^{-1}x^*\| \leqslant \|H^{-1}\|\ \|x^*\|$, 有

$$\langle H\bar{x},\bar{x}\rangle \geqslant \frac{\langle x^*,x_1\rangle^2}{\langle Hx_1,x_1\rangle} \geqslant \frac{\|x^*\|^2}{4\|H\|} \geqslant \frac{1}{4}\|H^{-1}\|^{-2}\|H\|^{-1}\|\bar{x}\|^2,$$

取 $\alpha := \frac{1}{4}\|H^{-1}\|^{-2}\|H\|^{-1}$, 即可得到期望的不等式.

现在证明条件是充分的. 因为 H 是一致正的, 二次规划问题 (4.336) 有强凸的目标函数. 引理 2.33 可推出, 对任何 $x^*\in X^*$, 问题 (4.336) 具有唯一的最优解 x, 由一阶最优性条件 $Hx=x^*$ 来刻画. 这表明 H 是一对一的映上的. 由开映射定理, 有 H^{-1} 是连续的, 因此 H 是可逆的. □

对等式约束问题中与最优性系统的线性化相对应的线性算子, 我们可推导类似的结果.

引理 4.124　设 X 与 Y 是 Banach 空间, $H\in\mathcal{L}(X,X^*)$ 是自伴随的, $A\in\mathcal{L}(X,Y)$. 设 H 在 $\operatorname{Ker}A$ 上是非负, 即

$$\langle Hx,x\rangle \geqslant 0, \quad \forall x\in\ker A.$$

则下述两个条件是等价的:

(i) 对任何 $(x^*,y)\in X^*\times Y$, 方程

$$Hx + A^*y^* = x^*, \quad Ax=y \tag{4.339}$$

具有唯一解 $(\bar{x},\bar{y}^*)\in X\times Y^*$.

(ii) 映射 A 是映上的, 且 H 在 $\ker A$ 上是一致正的, 即存在 $\alpha>0$ 满足

$$\langle Hx,x\rangle \geqslant \alpha\|x\|^2, \quad \forall x\in\ker A. \tag{4.340}$$

证明　我们证明由 (i) 可推出 (ii). 显然, 若对所有的 $(x^*,y)\in X^*\times Y$, (4.339) 具有唯一解, 则 A 是映上的. 因 H 在 $\ker A$ 上是非负的, 则 (4.339) 是凸优化问题

$$\min_{x\in X}\ \frac{1}{2}\langle Hx,x\rangle - \langle x^*,x\rangle, \quad \text{s. t. } Ax=y \tag{4.341}$$

的一阶最优性系统. 尤其, 对 $y=0$, 最优化问题

$$\min_{x\in X}\ \frac{1}{2}\langle Hx,x\rangle-\langle x^*,x\rangle,\quad \text{s. t. } x\in\ker A \tag{4.342}$$

具有唯一最优解. 视上述问题为 Banach 空间 $\ker A$ 上的无约束优化问题. 考虑空间 $\ker A$ 上的二次型 $Q_A(x):=\langle Hx,x\rangle$, 相联系的算子 $H_A:\ker A\to(\ker A)^*$ 定义为

$$\langle H_Ax,x'\rangle:=\frac{1}{2}[Q_A(x+x')-Q_A(x)-Q_A(x')].$$

算子 H_A 是自伴随的, 且对所有的 $x\in\ker A$ 有 $\langle H_Ax,x\rangle=\langle Hx,x\rangle$, 从而是非负的. 因为 (4.342) 有唯一解, 由 $H_Ax=x_A^*$ (其中 x_A^* 是 x^* 在 $\ker A$ 上的限制) 得, H_A 是一对一的且映上的. 则由引理 4.123 得到 (4.340).

现在证明由 (ii) 可推出 (i). 因为 A 是映上的, 存在 x_y 满足 $Ax_y=y$. 作变量变换 $x':=x-x_y$, 得到 (4.339) 具有唯一解 $(\bar{x},\bar{y}^*)$ 当且仅当 $y=0$ 时的相应系统有唯一解, 即

$$H\bar{x}+A^*\bar{y}^*=x^*,\quad \bar{x}\in\ker A.$$

这是当 $y=0$ 时问题 (4.341) 的最优性系统. 因为 A 是映上的, 目标函数在空间 $\ker A$ 上是凸的, 这一最优系统刻画了最优化问题 (4.342) 的解. 因为目标函数是严格凸的, 由引理 2.33 可推出解 $\bar{x}$ 的存在性和唯一性, 进而 $\bar{y}^*$ 的唯一性可得到, 因为 A 是映上的. □

研究如下的等式约束问题参数化族

$$(\mathrm{P}_u)\qquad \min_{x\in X}\ f(x,u)\quad \text{s. t. } G(x,u)=0,$$

其中 $f:X\times U\to\mathbb{R}$ 与 $G:X\times U\to Y$ 是二阶连续可微的, 相应的 Lagrange 函数是

$$L(x,\lambda,u):=f(x,u)+\langle\lambda,G(x,u)\rangle.$$

设 x_0 是非扰动问题 (P_{u_0}) 的局部最优解, $D_xG(x_0,u_0):X\to Y$ 是映上的, 则 x_0 与唯一的 Lagrange 乘子 λ_0 相联系, 且由命题 3.46, 下述二阶必要性条件成立①

$$D_{xx}^2L(x_0,\lambda_0,u_0)(h,h)\geqslant 0,\quad \forall h\in\ker[D_xG(x_0,u_0)]. \tag{4.343}$$

进一步, 由定理 3.70, 二阶增长条件在 x_0 处成立的充分必要条件是下述二阶充分条件成立：存在 $\alpha>0$, 满足

$$D_{xx}^2L(x_0,\lambda_0,u_0)(h,h)\geqslant\alpha\|h\|^2,\quad \forall h\in\ker[D_xG(x_0,u_0)]. \tag{4.344}$$

① 下面公式原著中为 $D_{xx}^2L(x_0,\lambda_0,u)$.

应用隐函数定理到最优性系统的主要困难在于刻画 Jacobi 算子的可逆性. 然而, 这是上述引理的直接结果.

定理 4.125　设 x_0 是 (非扰动) 问题 (P_{u_0}) 的局部最优解, λ_0 是相应的 Lagrange 乘子. 设 $D_xG(x_0,u_0)$ 是映上的 (因此 Lagrange 乘子 λ_0 是唯一的), 二阶充分条件 (4.344) 成立. 则下述结论成立:

(i) 存在 x_0 的邻域 $\mathcal{V}_X$, u_0 的邻域 $\mathcal{V}_U$, 满足对任何 $u \in \mathcal{V}_U$, 问题 (P_u) 的最优性系统

$$D_xL(x,\lambda,u)=0, \quad G(x,u)=0 \tag{4.345}$$

在 $\mathcal{V}_X \times Y^*$ 有唯一的最优解 $(\bar{x}(u),\overline{\lambda}(u))$.

(ii) 进一步, 可选取 $\mathcal{V}_X$ 与 $\mathcal{V}_U$ 满足 $\bar{x}(u)$ 的二阶增长条件成立, 即

$$f(x) \geqslant f(\bar{x}(u)) + c\|x-\bar{x}(u)\|^2, \quad \forall x \in \Phi(u) \cap \mathcal{V}_X,$$

其中 $c := \dfrac{1}{2}\alpha$.

(iii) 解映射 $(\bar{x}(\cdot),\overline{\lambda}(\cdot))$ 在 u_0 的邻域内是 Fréchet 可微的, 其微分 $(h,\mu) := (D\bar{x}(u_0)d, D\overline{\lambda}(u_0)d)$ 是下述线性化最优系统的最优解:

$$\begin{aligned}&D^2L(x_0,\lambda_0,u_0)(h,d) + D_xG(x_0,u_0)^*\mu = 0,\\&DG(x_0,u_0)(h,d)=0.\end{aligned} \tag{4.346}$$

证明　由引理 4.124, 应用隐函数定理 (见定理 5.14) 到 (x_0,λ_0,u_0) 邻域的最优性系统 (4.345). 得到 (i), 观察到, 由于 $D_xG(x_0,u_0)$ 是映上的, 因而 $D_xG(x_0,u_0)^*$ 是一对一的, 从而没有必要将 $\overline{\lambda}(u)$ 限定到 λ_0 的邻域内.

(ii) 假设结论不真. 则存在序列 $u_k \to u_0, x_k \to x_0$ 与 $\lambda_k \to \lambda_0$, 满足 (x_k,λ_k) 是最优性系统 (4.345) 的解, 存在 $z_k \to x_0$, $z_k \neq x_k$, $z_k \in \Phi(u_k)$ 对充分大的 k 成立, 满足

$$f(z_k,u_k) \leqslant f(x_k,u_k) + o(\|z_k-x_k\|^2). \tag{4.347}$$

令 $z_k = x_k + t_kh_k$, 其中 $t_k := \|z_k - x_k\|$, 从而 $\|h_k\| = 1$. 由

$$0 = G(z_k,u_k) = G(x_k,u_k) + t_kDG(x_k,u_k)h_k + o(t_k)$$

且 $DG(x_k,u_k) \to DG(x_0,u_0)$, 得到 $D_xG(x_0,u_0)h_k = o(1)$. 由开映射定理 (见 2.3.1 节), 存在 X 中的序列 $\{r_k\}$ 满足 $D_xG(x_0,u_0)r_k = D_xG(x_0,u_0)h_k$, $r_k = o(1)$. 则 $\hat{h}_k := h_k - r_k$ 满足 $D_xG(x_0,u_0)\hat{h}_k = 0$, $\|\hat{h}_k - h_k\| \mapsto 0$, 因此 $\|\hat{h}_k\| \to 1$. 因为

$D_xL(x_k,\lambda_k,u_k)=0$, 得

$$\begin{aligned} f(z_k,u_k)-f(x_k,u_k) &= L(z_k,\lambda_k,u_k)-L(x_k,\lambda_k,u_k) \\ &= \frac{1}{2}t_k^2D_{xx}^2L(x_k,\lambda_k,u_k)(h_k,h_k)+o(t_k^2) \\ &= \frac{1}{2}t_k^2D_{xx}^2L(x_0,\lambda_0,u_0)(\hat{h}_k,\hat{h}_k)+o(t_k^2) \\ &\geqslant \frac{1}{2}\alpha t_k^2\|\hat{h}_k\|^2+o(t_k^2) \\ &= \frac{1}{2}\alpha\|z_k-x_k\|^2+o(t_k^2), \end{aligned}$$

其中 $\alpha>0$ 由 (4.344) 给出. 得到矛盾.

最后, 结论 (iii) 是隐函数定理的一个直接结果. □

注 4.126 方程 (4.346) 可解释为下述 (二次规划)最优化问题的最优性系统

$$\begin{aligned} \min_{h\in X} \quad & D^2L(x_0,\lambda_0,u_0)((h,d),(h,d)), \\ \text{s. t.} \quad & DG(x_0,u_0)(h,d)=0. \end{aligned}$$

注 4.127 考虑等式与不等式约束的参数最优化问题

$$(\mathrm{P}'_u) \qquad \min_{x\in X} f(x,u) \quad \text{s. t.} \quad G(x,u)=0, \quad g_i(x,u)\leqslant 0, \quad i=1,\cdots,p,$$

其中 $f(x,u)$, $G(x,u)$ 与 $g_i(x,u)$ 是二次连续可微的. 令 x_0 是 (P'_{u_0}) 的局部最优解, 设映射

$$h\mapsto (D_xG(x_0,u_0)h,\{D_x\,g_i(x_0,u_0)h\}_{i\in I(x_0,u_0)})$$

是映上的, 其中 $I(x_0,u_0):=\{i:g_i(x_0,u_0)=0, i=1,\cdots,p\}$ 记起作用约束的指标集合. 设 λ_0 是与 x_0 相联系的唯一的 Lagrange 乘子. 进一步, 设严格互补松弛条件成立, 即对应于 (x_0,u_0) 处起作用的不等式约束的 λ_0 的所有分量是正的. 则对与 (x_0,u_0) 充分接近的所有 (x,u), Lagrange 乘子集合局部地为单点集且连续地依赖于 (x,u). 因此, 在 (x_0,u_0) 的邻域中, 起作用的不等式约束集合是不变的, 因此, (P'_u) 的 x_0 附近的局部最优解与下述等式约束问题的最优解是重合的

$$(\mathrm{P}''_u) \qquad \min_{x\in X} f(x,u) \quad \text{s. t.} \quad G(x,u)=0, \quad g_i(x,u)=0, \quad i\in I(x_0,u_0).$$

即在约束映射关于 x 的导数是映上的, 在严格互补条件成立的假设下, 局部分析可简化为等式约束问题的分析.

注 4.128 将局部扰动分析减化 (局部地) 为等式约束问题的扰动分析的另一情况是严格互补条件 (5.184) 成立的半定规划问题情况 (见 5.3.6 节结束的注记).

在 5.1 节将对基于方程的最优化问题的扰动分析方法做进一步的探讨.

4.9.2 最优值与最优解的一致近似

4.7 节已经证明, 在一定的正则性假设与某些形式的二阶充分条件下, 可以计算最优 (或 "接近最优") 解 $\bar{x}(u)$ 的方向导数. 有趣的是, 得到的沿形式为 $u_0+td+\frac{1}{2}t^2r$ 的抛物路径的最优解的一阶变化公式通过项 $D_uL(x_0,\lambda,u_0)r$ 显式地依赖于参数的二阶变化 $\frac{1}{2}t^2r$, $D_uL(x_0,\lambda,u_0)r$ 出现在相应的辅助的极小 - 极大问题中. 在 $\Lambda(x_0,u_0)$ 非单点集且约束映射依赖于 u 的情况下, 这往往导致方向导数 $\bar{x}'(u_0,d)$ 关于方向 d 的不连续性态. 例 4.99 清楚地说明了这一点. 在这样的情形下, Hadamard 或 Fréchet 意义下的方向可微性的较强形式是不成立的. 因此, 只有在一些特殊的情况, 如在 Lagrange 乘子唯一的情况或可行集不依赖于 u 的情况, 才可望能够导出 $\bar{x}(u)$ 的 Hadamard 或 Fréchet 方向可微性质.

先讨论可行集 Φ 不依赖于 u 的情况, 即最优值函数由形式 $v(u):=\inf\limits_{x\in\Phi}f(x,u)$ 给出. 设 Φ 是 X 的闭 (不必凸的) 子集, 函数 $f(x,u)$ 是二次连续可微的. 令 $x_0\in\mathcal{S}(u_0)$, 即 x_0 是 $f(\cdot,u_0)$ 在 Φ 上的极小点. 为简便起见, 设集合 Φ 在 x_0 处的内切锥与外切锥相同, 即 $T_\Phi(x_0)=T_\Phi^i(x_0)$.

由一阶必要性条件, 有

$$D_xf(x_0,u_0)h\geqslant 0,\quad h\in T_\Phi(x_0). \tag{4.348}$$

规定锥

$$C(x_0):=\{h\in T_\Phi(x_0):\ D_xf(x_0,u_0)h=0\} \tag{4.349}$$

是临界方向锥. 在可行集定义为约束集 $\Phi:=G^{-1}(K)$ 的形式, Robinson 约束规范成立时, 上述锥 $C(x_0)$ 与以前定义的临界锥是相同的. 令 $h\in C(x_0)$, 考虑下述形式的抛物路径 $x(t)$:

$$x(t):=x_0+th+\frac{1}{2}t^2w+o(t^2). \tag{4.350}$$

由内二阶切集的定义 3.28, 适当选取余项 $o(t^2)$, 对充分小的 $t>0$, 路径 $x(t)$ 可行的充要条件是 $w\in T_\Phi^{i,2}(x_0,h)$.

对给定的方向 $d\in U$, 考虑最优化问题

$$\min_{h\in C(x_0)}\{D^2f(x_0,u_0)((h,d),(h,d))-\sigma(-D_xf(x_0,u_0),T_\Phi^{i,2}(x_0,h))\}, \tag{4.351}$$

令 $\nu(d)$ 记其最优值. 只有当 $h\in T_\Phi(x_0)$ 时, $T_\Phi^{i,2}(x_0,h)$ 才可能非空. 又观察到, 对 $h=0$, 相应的二阶切集 $T_\Phi^{i,2}(x_0)$ 与 $T_\Phi(x_0)$ 重合, 因而是非空的. 进一步, 最优性条件 (4.348) 意味着 $-D_xf(x_0,u_0)\in[T_\Phi(x_0)]^-$. 所以, 对于 $h=0$, 上述问题中的 sigma 项是零, 因此有 $\nu(d)\leqslant D_{uu}^2f(x_0,u_0)(d,d)$.

命题 4.129 令 $x_0 \in \mathcal{S}(u_0)$, $\nu(d)$ 是问题 (4.351) 的最优值. 则

$$\limsup_{\substack{t\downarrow 0\\ d'\to d}} \frac{v(u_0+td')-v(u_0)-tD_u f(x_0,u_0)d'}{\frac{1}{2}t^2} \leqslant \nu(d). \tag{4.352}$$

证明 考虑点 $h \in C(x_0)$ 与点 $w \in T_\Phi^{i,2}(x_0,h)$. 存在 (4.350) 形式的路径 $x(t)$ 满足 $x(t) \in \Phi$, 因此, 对充分小的 $t>0$,

$$v(u_0+td') \leqslant f(x(t),u_0+td').$$

由二阶 Taylor 展式及 $D_x f(x_0,u_0)h=0$, 得到

$$\begin{aligned} f(x(t),u_0+td') = {} & f(x_0,u_0)+tD_u f(x_0,u_0)d' \\ & + \frac{1}{2}t^2[D_x f(x_0,u_0)w + D^2 f(x_0,u_0)((h,d),(h,d))] + o(t^2). \end{aligned}$$

结果, 对任何 $h \in C(x_0)$ 及 $w \in T_\Phi^{i,2}(x_0,h)$, (4.352) 左端的上极限小于或等于

$$D_x f(x_0,u_0)w + D^2 f(x_0,u_0)((h,d),(h,d)).$$

对这样的 h 与 w 取极小, 得到不等式 (4.352). □

注 4.130 注意到, 若最优值 $\nu(d)$ 是有限的, 上述命题的结果可以以下述的等价形式描述: 对任何满足当 $t\downarrow 0$, $tr(t)\to 0$ 的路径 $r(\cdot):\mathbb{R}_+\to U$ 与 $u(t):=u_0+td+\frac{1}{2}t^2 r(t)$, 下述不等式成立:

$$v(u(t)) \leqslant v(u_0)+tD_u f(x_0,u_0)d+\frac{1}{2}t^2[D_u f(x_0,u_0)r(t)+\nu(d)]+o(t^2). \tag{4.353}$$

在 (4.352) 的左端置 $d'=d+\frac{1}{2}tr(t)$ 就得到上式.

注 4.131 考虑最大值函数 $\bar{v}(u):=\sup_{x\in\Phi} f(x,u)$. 由上述命题, 对任何 $x_0 \in \operatorname{argmax}_{x\in\Phi} f(x,u_0)$,

$$\liminf_{\substack{t\downarrow 0\\ d'\to d}} \frac{\bar{v}(u_0+td')-\bar{v}(u_0)-tD_u f(x_0,u_0)d'}{\frac{1}{2}t^2} \geqslant \bar{\nu}(d), \tag{4.354}$$

其中 $\bar{\nu}(d)$ 是下述问题的最优值

$$\max_{h\in C(x_0)} \{D^2 f(x_0,u_0)((h,d),(h,d))+\sigma(D_x f(x_0,u_0),T_\Phi^{i,2}(x_0,h))\}. \tag{4.355}$$

不等式 (4.354) 可以表示为下述形式

$$d^2\bar{v}(u_0|D_u f(x_0,u_0))(d) \geqslant \bar{\nu}(d), \tag{4.356}$$

其中 $\bar{v}(\cdot)$ 的二阶次导数由 (3.183) 定义.

注 4.132　设可行集由集合包含约束定义, 即 $\Phi := G^{-1}(K)$, 其中映射 $G: X \to Y$ 是二次连续可微的, K 是 Y 的闭凸子集. 进一步, 设 Robinson 约束规范

$$0 \in \operatorname{int}\{G(x_0) + DG(x_0)X - K\} \tag{4.357}$$

成立. 此种情形, Lagrange 乘子集合 $\Lambda(x_0, u_0)$ 是非空的有界的. 此时, 由链式法则 (见命题 3.33) 有

$$T_\Phi^{i,2}(x_0, h) = DG(x_0)^{-1}[T_K^{i,2}(G(x_0), DG(x_0)h) - D^2G(x_0(h,h)].$$

在此种情况下, 最优化问题 (4.351) 的目标函数可以表示为下述最优化问题

$$\begin{array}{ll}\min\limits_{w \in X} & D_x f(x_0, u_0)w + D^2 f(x_0, u_0)((h,d),(h,d)) \\ \text{s. t.} & DG(x_0)w + D^2G(x_0)(h,h) \in \mathcal{T}_K^2(h),\end{array}$$

其中

$$\mathcal{T}_K^2(h) := T_K^{i,2}(G(x_0), DG(x_0)h). \tag{4.358}$$

上述问题的对偶 (见 (4.208) 的推导) 是

$$\max_{\lambda \in \Lambda(x_0,u_0)} \{D^2L(x_0, \lambda, u_0)((h,d),(h,d)) - \sigma(\lambda, \mathcal{T}_K^2(h))\}.$$

在 Robinson 约束规范 (4.357) 下, 上述互为对偶的问题间没有对偶间隙, 因而 $\nu(d)$ 是下述极小 - 极大问题的最优值

$$\min_{h \in C(x_0)} \max_{\lambda \in \Lambda(x_0,u_0)} \{D^2L(x_0, \lambda, u_0)((h,d),(h,d)) - \sigma(\lambda, \mathcal{T}_K^2(h))\}. \tag{4.359}$$

注意, 如前面所指出的, 若 K 是多面体, 则 (4.395) 中的 sigma 项消失.

在当前情形下, 约束映射与参数 u 是独立的, 上界 (4.354) 与 (4.353) 在某种意义下是一致的, 因此强于命题 4.83 的上界结果.

定理 4.133　设

(i) 空间 X 是有限维的.

(ii) 对每一 $h \in C(x_0)$, 集合 Φ 在点 x_0 处沿方向 h 是二阶正则的.

(iii) 存在 (P_u) 的 $o(\|u-u_0\|^2)$ 最优解 $\bar{x}(u)$, 当 $u \to u_0$ 时, 它以 Lipschitz 率收敛到 x_0, 即 $\|\bar{x}(u) - x_0\| = O(\|u - u_0\|)$.

则

$$\lim_{\substack{t \downarrow 0 \\ d' \to d}} \frac{v(u_0 + td') - v(u_0) - tD_u f(x_0, u_0)d'}{\frac{1}{2}t^2} = \nu(d). \tag{4.360}$$

若进一步, 问题 (4.351) 具有唯一的最优解 $\hat{h} = \hat{h}(d)$, 则 $\bar{x}(u)$ 在 u_0 处沿方向 d 是方向可微的, 而且是 Hadamard 意义下方向可微的, 且 $\bar{x}'(u_0, d) = \hat{h}(d)$.

证明 最优值函数所需的上方估计由 (4.352) 得到, 相应的下方估计由类似于 4.7.3 节中的定理 4.100 与 4.101 证明中的推导得到. 我们简要地给出相应的步骤.

考虑序列 $t_n \downarrow 0$, $d_n \to d$, 定义

$$u_n := u_0 + t_n d_n, \quad x_n := \bar{x}(u_n), \quad h_n := t_n^{-1}(x_n - x_0).$$

因为 $\bar{x}(u)$ 是 Lipschitz 稳定的, 则 h_n 是有界的. 因此, 由于空间 X 是有限维的, 如有必要可选取一子列, 不妨设 h_n 收敛到向量 h. 可以证明 $h \in C(x_0)$. 因此, 可以表示为 $x_n = x_0 + t_n h + \frac{1}{2}t_n^2 w_n$, 其中 $t_n w_n \to 0$. 由集合 Φ 在 x_0 处沿方向 h 的二阶正则性有 $\operatorname{dist}(w_n, T_\Phi^{i,2}(x_0, h)) \to 0$. 由于 $v(u_n) = f(x_n, u_n) + o(t_n^2)$, 用 $f(x_n, u_n)$ 相应的 Taylor 展示可完成定理的证明. □

注 4.134 由 (4.360), 若最优值 $\nu(d)$ 是有限的, 则

$$v''(u_0; d, r) = D_u f(x_0, u_0) r + \nu(d). \tag{4.361}$$

进一步, 对任何满足 $t \downarrow 0$ 时 $tr(t) \to 0$ 的路径 $u(t) := u_0 + td + \frac{1}{2}t^2 r(t)$, $t \geqslant 0$, 展式

$$v(u(t)) = v(u_0) + tD_u f(x_0, u_0) d + \frac{1}{2}t^2[D_u f(x_0, u_0) r(t) + \nu(d)] + o(t^2) \tag{4.362}$$

成立, 因此, 函数 $v(\cdot)$ 与 $-v(\cdot)$ 在 u_0 处沿方向 d 均是二阶上图正则的 (见定义 3.94).

如果集合 $\mathcal{S}(u_0)$ 是有限的, 如 $\mathcal{S}(u_0) = \{x_1, \cdots, x_k\}$ 且在每一点 x_i, $i = 1, \cdots, k$ 处二阶增长条件成立, 下确界–紧致性条件满足, 则 $\mathcal{S}(u)$ 在 u_0 处是 Lipschitz 稳定的 (见命题 4.36). 若 X 还是有限维的, 集合 Φ 在每一点 x_i 处是二阶正则的, $i = 1, \cdots, k$, 则

$$v''(u_0; d, r) = \min_{i \in \mathcal{I}(d)} \{D_u f(x_i, u_0) r + \nu_i(d)\}, \tag{4.363}$$

其中 $\mathcal{I}(d) := \underset{i \in \{1, \cdots, k\}}{\operatorname{argmin}} D_u f(x_i, u_0) d$, $\nu_i(d)$ 是问题 (4.351) 将 x_0 替换为 x_i 得到的问题的最优值. 此种情况, 函数 $v(\cdot)$ 与 $-v(\cdot)$ 在 u_0 处是二阶上图正则的.

注 4.135 若可行集 Φ 由集合包含约束定义, 如注 4.132, 且 Robinson 约束规范成立, 则 $\nu(d)$ 由极小–极大问题 (4.359) 的最优值给出. 进一步, 此种情况, 若相应的集合 $K \subset Y$ 在 $G(x_0)$ 处沿方向 $DG(x_0)h$ 关于映射 $DG(x_0)$ 是二阶正则的, 则集合 Φ 在 x_0 处沿方向 h 是二阶正则的 (见命题 3.88). 此种情况, sigma 项 $\sigma(\lambda, \mathcal{T}_K^2(h))$ 小于或等于零, 则在上述定理的假设下, 最优值 $\nu(d)$ 是有限的, 相应的最优化问题有一最优解 $\hat{h} = \hat{h}(d)$(可能不是唯一的).

下面讨论可行集由等式与有限个不等式约束定义的情况, 即 $Y := Y_1 \times \mathbb{R}^m$, Y_1 是 Banach 空间, $K = \{0\} \times \mathbb{R}^m_- \subset Y$, $\Phi(u) := \{x : G(x,u) \in K\}$. 设 Robinson 约束规范关于简约锥 K_0 成立, 即严格约束规范 (4.119) 成立. 由命题 4.47 可推出 $\Lambda(x_0,u_0) = \{\lambda_0\}$, 若空间 Y_1 是有限的, 即可行集由有限多个约束定义, 则 λ_0 的唯一性等价于条件 (4.119).

为便于记号, 设在点 (x_0,u_0) 处所有的不等式约束均是起作用的, 即 $G(x_0,u_0)$ 是空间 Y 的零向量, $\lambda_0 = (\mu, \lambda_1, \cdots, \lambda_m)$, $\mu \in Y_1^*$, $\lambda_i > 0, i = 1, \cdots, m_1; \lambda_i = 0, i = m_1 + 1, \cdots, m_0$, 则简约锥 K_0 可以表示为下述形式

$$K_0 = \{0\} \times \{0\} \times \mathbb{R}^{m-m_1}_- \subset Y_1 \times \mathbb{R}^{m_1} \times \mathbb{R}^{m-m_1}, \tag{4.364}$$

因此, 对应于非零 Lagrange 乘子的不等式约束被处理为等式约束时, 条件 (4.119) 变为 Mangasarian-Fromovitz 约束规范. 注意, 上述情形, 由 K_0 生成的线性空间可以表示为

$$\mathrm{Sp}\ (K_0) = \{0\} \times \{0\} \times \mathbb{R}^{m-m_1},$$

且临界锥为

$$C(x_0) = \{h \in X : D_xG(x_0,u_0)h \in K_0\}. \tag{4.365}$$

设 $f(\cdot,u_0)$, $G(\cdot,u_0)$ 是二次连续可微的, $f(x_0,\cdot)$, $D_xf(x_0,\cdot)$, $G(x_0,\cdot)$ 与 $D_xG(x_0,\cdot)$ 在 u_0 附近是 Lipschitz 连续的, 则非扰动问题 (P_{u_0}) 在 x_0 处的二阶增长条件成立当且仅当对某一 $\beta > 0$, 下述二阶条件成立

$$D^2_{xx}L(x_0,\lambda_0,u_0)(h,h) \geqslant \beta\|h\|^2, \quad \forall h \in C(x_0) \tag{4.366}$$

(见定理 3.70).

考虑下述参数化问题

$$(\mathrm{F}_u) \qquad \begin{aligned} \min_{h\in X} \quad & D_xL(x_0,\lambda_0,u)h + \frac{1}{2}D^2_{xx}L(x_0,\lambda_0,u_0)(h,h), \\ \text{s. t.} \quad & G(x_0,u) + D_xG(x_0,u_0)h \in K_0. \end{aligned} \tag{4.367}$$

我们也用下述二阶条件: 存在 $\beta > 0$, 满足

$$D^2_{xx}L(x_0,\lambda_0,u_0)(h,h) \geqslant \beta\|h\|^2, \quad \forall h \in \mathcal{L}, \tag{4.368}$$

其中

$$\mathcal{L} := \{h \in X : D_xG(x_0,u_0)h \in \mathrm{Sp}\ (K_0)\}.$$

因为 $C(x_0) \subset \mathcal{L}$, 上述条件比二阶条件 (4.366) 要强.

做如下的观察. 由相对于锥 K_0 的 Robinson 约束规范成立可得, 对充分接近 u_0 的任何 u, 问题 (F_u) 的可行集是非空的. 这一可行集包含在下述仿射空间中,

$$\{h \in X : G(x_0, u) + D_x G(x_0, u_0)h \in \mathrm{Sp}\ (K_0)\},$$

对应于这一仿射空间的线性空间与线性空间 $\mathcal{L}$ 是重合的. 由 (4.368), 线性空间 $\mathcal{L}$ 赋予 X 的范数所引导的范数是可 Hilbert 化的. 进一步, 在 (4.368) 之下, 问题 (F_u) 简约到上述仿射空间具有强凸的费用函数, 因而具有唯一的最优解, 后面把它记为 $\bar{h}(u)$. 因为 $D_x L(x_0, \lambda_0, u_0) = 0$, $G(x_0, u_0) \in K$, 由二阶条件 (4.368) 可推出 $\bar{h}(u_0) = 0$, 对 $u = u_0$, 问题 (F_{u_0}) 具有零的 Lagrange 乘子对应最优解 $\bar{h}(u_0) = 0$, 对问题 (F_{u_0}), Robinson 约束规范与问题 (P_{u_0}) 相对于锥 K_0 的 Robinson 约束规范是相同的. 结合二阶条件 (4.368), 由定理 4.51 与命题 4.47 可推出 $\bar{h}(u) = O(\|u - u_0\|)$.

定理 4.136 设 (P_u) 的可行集由等式和有限多个不等式约束来定义, 令 $\bar{x}(u)$ 是 (P_u) 的 $\varepsilon(u)$ 最优解, 它当 $u \to u_0$ 时收敛到 $x_0 \in \mathcal{S}(u_0)$, $\varepsilon(u) = o(\|u - u_0\|^2)$. 设

(i) 函数 $f(\cdot, u)$ 与 $G(\cdot, u)$ 是二次连续可微的, $D^2_{xx} f(x, u)$, $D^2_{xx} G(x, u)$ 在 (x_0, u_0) 处是连续的, $f(x_0, \cdot)$, $D_x f(x_0, \cdot)$, $G(x_0, \cdot)$, $D_x G(x_0, \cdot)$ 在 u_0 附近是 Lipschitz 连续的.

(ii) Robinson 约束规范相对于简约锥 K_0 成立 (即严格约束规范成立).

(iii) 二阶条件 (4.368) 成立.

则

$$\|\bar{x}(u) - x_0 - \bar{h}(u)\| = o(\|u - u_0\|), \tag{4.369}$$

其中 $\bar{h}(u)$ 是问题 (F_u) 的最优解.

证明 为了记号的方便, 设所有的不等式约束在点 (x_0, u_0) 处是起作用的, $\lambda_0 = (\mu, \lambda_1, \cdots, \lambda_m)$, $\mu \in Y_1^*$ 且 $\lambda_i > 0, i = 1, \cdots, m_1$; $\lambda_i = 0, i = m_1 + 1, \cdots, m$. 根据定理 4.51 与命题 4.47, 由假设 (i)~(iii) 可得, $\bar{x}(u)$ 在 x_0 处是 Lipschitz 稳定的且 $\bar{h}(u) = O(\|u - u_0\|)$.

由假设 (ii), 关于 K 在 (x_0, u_0) 处的 Robinson 约束规范成立, 由 Ekeland 变分原理, 对所有充分接近 u_0 的 u, 存在 (P_u) 的另一个 $\varepsilon(u)$ 的最优解 $\hat{x}(u)$, 满足 $f(\hat{x}(u), u) \leqslant f(\bar{x}(u), u)$, $\|\hat{x}(u) - \bar{x}(u)\| = \varepsilon(u)^{\frac{1}{2}} = o(\|u - u_0\|)$, 在这一点处一阶最优性 (或 "近似最优性") 条件 (3.32) 成立. 令 $\hat{\lambda}(u) = (\hat{u}, \hat{\lambda}_1, \cdots, \hat{\lambda}_m)$ 是相应的 Lagrange 乘子, 有 $\hat{\lambda}_i g_i(\hat{x}(u), u) = 0, i = 1, \cdots, m$, 其中 g_i 是不等式约束的约束函数. 由连续性推证得 $\hat{\lambda}_i > 0, i = 1, \cdots, m_1$, 从而对充分接近于 u_0 的所有 u, 所有的不等式约束 $g_i, i = 1, \cdots, m_1$ 在 $(\hat{x}(u), u)$ 处是起作用的. 所以, 对接近于 u_0 的所有 u, $\hat{x}(u)$ 是下述问题的 $\varepsilon(u)$ 最优解:

$$(\mathrm{P}'_u) \qquad \min_{x \in X} L(x, \lambda_0, u) \quad \text{s.t.} \quad G(x, u) \in K_0. \tag{4.370}$$

我们现在证明, 根据命题 4.37 的估计, 问题 (P'_u) 与 (F_u) 彼此是相互充分接近的. 注意到, 由二阶条件 (4.368), 问题 (F_u) 在 $\bar{h}(u)$ 处的二阶增长条件成立, 相应的常数与 u 无关. 不失一般性, 可设 $x_0 = 0$.

首先, 在 0 的包含 $\hat{x}(u)$ 与 $\bar{h}(u)$ 的邻域 N 上, 估计函数 (问题 (P'_u) 与 (F_u) 的目标函数的差函数)

$$H(x,u) := L(x,\lambda_0,u) - D_xL(x_0,\lambda_0,u)x - \frac{1}{2}D^2_{xx}L(x_0,\lambda_0,u_0)(x,x)$$

的 Lipschitz 常数. 注意, 由于 $\hat{x}(u)$ 与 $\bar{h}(u)$ 是 Lipschitz 稳定的, 可取 $N = B(0,r)$, 其中半径 r 是 $O(\|u-u_0\|)$ 阶的. 由于

$$D_xH(x,u) = D_xL(x,\lambda_0,u) - D_xL(x_0,\lambda_0,u) - D^2_{xx}L(x_0,\lambda_0,u_0)x,$$

$D^2_{xx}L(\cdot,\lambda_0,\cdot)$ 在 (x_0,u_0) 处是连续的, 得到差函数 $H(\cdot,u)$ 的 Lipschitz 常数可取为 $o(\|u-u_0\|)$ 阶的. 由稳定性定理, 问题 (P'_u) 与 (F_u) 限定在邻域 N 上的可行集间的 Hausdorff 距离是 $o(\|u-u_0\|)$ 阶的. 最后, 因为 $D_xL(x_0,\lambda_0,u_0)=0$, (P'_u) 与 (F_u) 在邻域 N 上的目标函数的 Lipschitz 常数是 $O(\|u-u_0\|)$ 阶的. 由估计 (4.96) 可得

$$\|\hat{x}(u) - \bar{h}(u)\| = o(\|u-u_0\|),$$

因此 (4.369) 成立. □

函数 $f(x,u)$ 与 $G(x,u)$ 关于联合变量 x 与 u 为二次连续可微的情况也可以对所涉及的函数关于 u 展开. 考虑近似问题

$$(\mathrm{F}^*_d) \qquad \begin{array}{ll} \min\limits_{h\in X} & D^2_{xu}L(x_0,\lambda_0,u_0)(h,d) + \frac{1}{2}D^2_{xx}L(x_0,\lambda_0,u_0)(h,h), \\ \text{s. t.} & D_xG(x_0,u_0)h + D_uG(x_0,u_0)d \in T_{K_0}(G(x_0,u_0)). \end{array}$$

用 $h^*(d)$ 记问题 (F^*_d) 的最优解. 在定理 4.136 的假设下, 有 $h^*(d)$ 存在且唯一, 并且

$$\|\bar{x}(u_0+d) - x_0 - h^*(d)\| = o(\|d\|). \tag{4.371}$$

进一步, $h^*(\cdot)$ 是正齐次的. 所以在这种情况下, $\bar{x}(u)$ 是 Fréchet 意义下方向可微的, 且 $\bar{x}'(u_0,d) = h^*(d)$. 注意到, 问题 (F^*_d) 的可行集由下述的约束定义:

$$\begin{aligned} &DG(x_0,u_0)(h,d) \in T_K(G(x_0,u_0)), \\ &\langle\lambda_0, DG(x_0,u_0)(h,d)\rangle = 0. \end{aligned} \tag{4.372}$$

这些约束意味着 $h \in \mathcal{S}(\mathrm{PL}_d)$ 与 $\mathcal{S}(\mathrm{DL}_d) = \{\lambda_0\}$(见 (4.47)), 即 (F^*_d) 的可行集与 $\mathcal{S}(\mathrm{PL}_d)$ 重合. 我们得到下述结果.

命题 4.137 令 (P_u) 的可行集由等式与有限多个不等式约束定义, 设 $\bar{x}(u)$ 是当 $u \to u_0$ 时收敛到 $x_0 \in \mathcal{S}(u_0)$ 的问题 (P_u) 的 $\varepsilon(u)$ 最优解, $\varepsilon(u) = o(\|u-u_0\|^2)$. 设

(i) 函数 $f(\cdot,\cdot)$ 与 $G(\cdot,\cdot)$ 是二次连续可微的.

(ii) 相对于简约锥 K_0, Robinson 约束规范成立 (即严格约束规范成立), 且

(iii) 二阶条件 (4.368) 成立.

则 $\bar{x}(u)$ 在 u_0 处是 Fréchet 意义下方向可微的, $\bar{x}'(u_0,d)$ 与下述问题的最优解重合,

$$\min_{h\in\mathcal{S}(\mathrm{PL}_d)} \quad D^2_{xu}L(x_0,\lambda_0,u_0)(h,d) + \frac{1}{2}D^2_{xx}L(x_0,\lambda_0,u_0)(h,h). \tag{4.373}$$

注 4.138 注意到, 关于锥 K_0 的 Robinson 约束规范 (即严格约束规范) 在下述定义下是非稳定的, 即数据的任意小的扰动会破坏它. 如考虑这样一个情况, 具有有限多个不等式约束, 满足 $D_x f(x_0,u_0)=0$. 此种情况 $K_0=K$, 相对于 K_0 的 Robinson 约束规范成为 Mangasarian-Fromovitz 约束规范. 若在 (x_0,u_0) 处起作用的不等式约束的梯度是线性相关的, 则使 $D_x f(\bar{x}(u),u)\neq 0$ 任意小的扰动将导致相应的 Lagrange 乘子是不唯一的. 此种情形, 即使在最强的二阶条件下, 在 u_0 的任何邻域, $\bar{x}(u)$ 可能不是 Lipschitz 连续的. 所以, 定理 4.136 的假设不能保证 $\bar{x}(u)$ 在 u_0 附近的 Lipschitz 连续性, 尽管在这些条件下 $\bar{x}(u)$ 是 Lipschitz 稳定的. 最优解的 Lipschitz 连续性将在 5.1 节进一步讨论.

下述定理导出最优值函数的 “一致的” 二阶展式.

定理 4.139 设 (P_u) 的可行集由等式和有限多个不等式约束定义. 设

(i) 存在 (P_u) 的一个 $o(\|u-u_0\|^2)$ 最优解 $\bar{x}(u)$, 当 $u\to u_0$ 时, 它收敛到 $x_0 \in \mathcal{S}(u_0)$.

(ii) 函数 $f(\cdot,\cdot)$ 与 $G(\cdot,\cdot)$ 是二次连续可微的.

(iii) Robinson 约束规范相对于简约锥 K_0 成立 (即严格的约束规范成立), 且

(iv) 二阶条件 (4.366) 成立.

则

$$\begin{aligned} v(u_0+d) = {} & v(u_0) + D_u L(x_0,\lambda_0,u_0)d \\ & + \frac{1}{2}\inf_{h\in\mathcal{S}(\mathrm{PL}_d)} D^2L(x_0,\lambda_0,u_0)((h,d),(h,d)) + o(\|d\|^2), \end{aligned} \tag{4.374}$$

其中 $\mathcal{S}(\mathrm{PL}_d)$ 是满足约束 (4.372) 的 h 的集合, 即相应的线性化问题的最优解集.

证明 如定理 4.136 的证明, 对充分接近于 u_0 的所有的 u, 问题 (P_u) 与 (P'_u)((P'_u) 的定义见 (4.370)) 具有同一 $o(\|u-u_0\|^2)$ 最优解. 结果, (P_u) 与 (P'_u) 最优值之差是 $o(\|u-u_0\|^2)$ 阶的, 因此, 只需要对问题 (P'_u) 的最优值推导出展开式 (4.374). 进一步, 在二阶条件 (4.366) 下, $\hat{x}(u)$ 在 u_0 处是 Lipschitz 稳定的.

对给定的 d, 令 h 是问题 (F_d^*) 的可行点, 即 $h \in \mathcal{S}(\mathrm{PL}_d)$, 则由于问题 (P'_{u_0}) 的 Robinson 约束规范成立, 由稳定性定理, 存在 (P'_u) 的可行点 $\tilde{x}$, 其中 $u = u_0 + d$, 这一可行解满足 $\|x_0 + h - \tilde{x}\| = o(\|h\| + \|d\|)$, 有 $v(u) \leqslant L(\tilde{x}, \lambda_0, u)$, 且注意到 $D_x L(x_0, \lambda_0, u_0) = 0$, 由 $L(\tilde{x}, \lambda_0, u)$ 的二阶 Taylor 展式得到

$$v(u) \leqslant v(u_0) + D_u L(x_0, \lambda_0, u_0)d + \frac{1}{2}D^2 L(x_0, \lambda_0, u_0)((h, d), (h, d)) + o(\|h\|^2 + \|d\|^2).$$

因为 h 是 $\mathcal{S}(\mathrm{PL}_d)$ 的任意点, 由于二阶条件 (4.366) 成立, 只需考虑在 (4.374) 右端优化问题中 h 是 $O(\|d\|)$ 阶的, 得到 $v(u)$ 小于或等于 (4.374) 的右端.

由 $\hat{x}(u)$ 是问题 (P'_u) 的 $o(\|u - u_0\|^2)$ 最优 (且 Lipschitz 稳定) 解, 可以得到另一方向的不等式. □

于是, 在上述定理的假设之下, 最优值函数 $v(u)$ 在 u_0 处是 Fréchet 可微的且 $Dv(u_0) = D_u L(x_0, \lambda_0, u_0)$. 进一步, $v(u)$ 在 u_0 处是二阶方向可微的, 满足

$$v''(u_0; d, r) = D_u L(x_0, \lambda_0, u_0)r + \inf_{h \in \mathcal{S}(\mathrm{PL}_d)} D^2 L(x_0, \lambda_0, u_0)((h, d), (h, d)), \tag{4.375}$$

且对满足 $t \downarrow 0$ 时 $tr(t) \to 0$ 的任何路径 $r(\cdot) : \mathbb{R}_+ \to U$, 有

$$v\left(u_0 + td + \frac{1}{2}t^2 r(t)\right) = v(u_0) + tDv(u_0)d + \frac{1}{2}t^2 v''(u_0; d, r(t)) + o(\|t\|^2). \tag{4.376}$$

4.9.3 非孤立最优点的二阶分析

这一节讨论非扰动问题的最优解集 $\mathcal{S}(u_0)$ 不必是单点集的情况. 若 $\mathcal{S}(u_0)$ 是有限的, 则前面几节讨论的二阶分析可以应用在每一 $\mathcal{S}(u_0)$ 的点上, 因此这样的情况从本质上可用已经建立的方法来处理. 然而, 若 $\mathcal{S}(u_0)$ 具有连通部分时, 就需要更深入的分析, 只有部分结果是已得到的.

当然, 不管 $\mathcal{S}(u_0)$ 是否是有限的, (在适当的假设下) 最优值函数 $v(u)$ 的上方方向估计 (4.205) 与 (4.217) 可应用于 $\mathcal{S}(u_0)$ 的每一点 x_0, 下方方向估计的导出则是十分微妙的. 这一节, 设空间 X 是有限维的向量空间 (赋予欧氏范数), 最优解集 $\mathcal{S}(u_0)$ 是非空的, 考虑一种特殊情形, 其中 $\mathcal{S}(u_0)$ 是可行集 $\Phi(u_0)$ 的光滑子流形. 假设 (P_u) 的可行集 $\Phi(u)$ 由有限多个约束定义:

$$\Phi(u) := \{x : g_i(x, u) = 0,\ i = 1, \cdots, q;\ g_i(x, u) \leqslant 0,\ i = q + 1, \cdots, p\}, \tag{4.377}$$

其中约束函数 $g_i(x, u)$, $i = 1, \cdots, p$ 与目标函数 $f(x, u)$ 是二阶连续可微的.

假设在给定的最优点 $x_0 \in \mathcal{S}(u_0)$ 处线性无关的约束规范成立. 即向量 $D_x g_i(x_0, u_0)$, $i = \{1, \cdots, q\} \cup I(x_0, u_0)$ 是线性无关的, 其中

$$I(x, u) := \{i : g_i(x, u) = 0,\ i = q + 1, \cdots, p\}.$$

由线性无关的约束规范可得存在唯一的 Lagrange 乘子向量 $\lambda(x_0)$. 由定理 4.26 得到下述结论.

命题 4.140 *设线性无关的约束规范在每一点 $x_0 \in \mathcal{S}(u_0)$ 处均成立且下确界–紧致性条件成立, 则最优值函数 $v(u)$ 在 u_0 处是 Hadamard 方向可微的, 且*

$$v'(u_0, d) = \inf_{x \in \mathcal{S}(u_0)} D_u L(x, \lambda(x), u_0)d. \tag{4.378}$$

进一步, 问题 $(\mathrm{P}_{u(t)})$(其中 $u(t) := u_0 + td + o(t)$) 的最优解路径 $\bar{x}(t)$ 当 $t \downarrow 0$ 时的每一极限点均包含在下述集合中

$$\mathcal{S}_1(x_0, d) := \operatorname*{argmin}_{x \in \mathcal{S}(u_0)} D_u L(x, \lambda(x), u_0)d. \tag{4.379}$$

若线性无关的约束规范在点 $x_0 \in \mathcal{S}(u_0)$ 处成立, 则由隐函数存在定理, 集合

$$\Theta(x_0, u_0) := \{x : g_i(x, u_0) = 0, i \in \{1, \cdots, q\} \cup I(x_0, u_0)\} \tag{4.380}$$

是点 x_0 邻域内的光滑 (可微) 流形. 由连续性可得, 对充分接近于 x_0 的所有的 x, $g_i(x, u_0) \leqslant 0$, $i = \{q+1, \cdots, p\} \setminus I(x_0, u_0)$, 因此, 限定在 x_0 的邻域上, $\Theta(x_0, u_0)$ 是 $\Phi(u_0)$ 的子集. 进一步假设 $\mathcal{S}(u_0)$ 是 x_0 邻域内 $\Theta(x_0, u_0)$ 的光滑子流形. 分别用 $T_{\mathcal{S}}(x_0)$ 与 $T_\Theta(x_0)$ 记 $\mathcal{S}(u_0)$ 与 $\Theta(x_0, u_0)$ 在 x_0 处的切空间.

注意到

$$T_\Theta(x_0) = \{h : D_x g_i(x_0, u_0)h = 0, i = \{1, \cdots, q\} \cup I(x_0, u_0)\},$$

对于非起作用约束, 即对 $i \in \{q+1, \cdots, p\} \setminus I(x_0, u_0)$, 相应的 Lagrange 乘子 $\lambda_i(x_0)$ 是零, 严格互补条件意味着 $\lambda_i(x_0) > 0$, $i \in I(x_0, u_0)$.

命题 4.141 *设线性无关的约束规范与严格互补条件在点 $x_0 \in \mathcal{S}(u_0)$ 成立且 $\mathcal{S}(u_0)$ 在 x_0 的邻域是 $\Theta(x_0, u_0)$ 的光滑的子流形, 则对非扰动问题 (P_{u_0}), 在 x_0 的邻域内二阶增长条件成立的充分必要条件是下述二阶条件成立:*

$$D^2_{xx} L(x_0, \lambda(x_0), u_0)(h, h) > 0, \quad \forall h \in [T_{\mathcal{S}}(x_0)]^\perp \cap T_\Theta(x_0), \quad h \neq 0. \tag{4.381}$$

证明 由线性无关约束规范与严格互补条件可推出临界锥 $C(x_0)$ 与切空间 $T_\Theta(x_0)$ 重合.

设二阶增长条件在 x_0 处成立, 令 $h \in [T_{\mathcal{S}}(x_0)]^\perp \cap T_\Theta(x_0)$. 由于 $h \in T_\Theta(x_0)$ 与线性无关约束规范成立, 存在对问题 (P_{u_0}) 可行的路径 $x(t) = x_0 + th + o(t)$. 因为 $\mathcal{S}(u_0)$ 是光滑的子流形, 不难验证, h 是 $\mathcal{S}(u_0)$ 在 x_0 处的追近法向量 (见定义 3.142),

因此 $\operatorname{dist}(x(t), \mathcal{S}(u_0)) = t\|h\| + o(t)$. 因为 $f(x(t), u_0) = L(x(t), \lambda(x_0), u_0)$, 由二阶增长条件得到

$$L(x_n, \lambda(x_0), u_0) \geqslant L(\tilde{x}_n, \lambda(x_0), u_0) + c\|\tilde{x}_n - x_n\|^2 \tag{4.382}$$

对所有的充分大的 n 成立, $c > 0$ 是某一常数. 因为 $D_x L(x_0, \lambda(x_0), u_0) = 0$, 由二阶 Taylor 展式并取极限可推出 $D_{xx}^2 L(x_0, \lambda(x_0), u_0)(h, h) > 0$, 这是需要证明的.

相反地, 设二阶条件 (4.381) 成立. 用反证法. 假设二阶增长条件不成立. 可推出存在可行点序列 $x_n \in \Phi(u_0)$ 收敛到 x_0, 满足①

$$f(x_n) \leqslant f(x_0) + o\Big(\operatorname{dist}(x_n, \mathcal{S}(u_0))^2\Big) \tag{4.383}$$

令 $\tilde{x}_n$ 是 $\mathcal{S}(u_0)$ 中的 (按欧氏范数) 与 x_n 最近的点, $h_n := (x_n - \tilde{x}_n)/\|x_n - \tilde{x}_n\|$. 如有必要, 可选取一子序列, 不妨设 h_n 收敛到一点 h. 因为 $\mathcal{S}(u_0)$②是 x_0 附近的光滑流形, 对充分大的 n, 有 h_n 于 $\tilde{x}_n$ 点处垂直于 $\mathcal{S}(u_0)$③, 因此 $h \in [T_{\mathcal{S}}(u_0)]^{\perp}$④. 由 (4.382) 有 $D_x f(\tilde{x}_n, u_0) h_n \leqslant o(1)$, 因此 $D_x f(x_0, u_0) h \leqslant 0$, 并且对所有的 $i \in I(x_0, u_0)$, 由严格互补条件, 对充分大的 n, $g_i(\tilde{x}_n, u_0) = 0$, 所以 $D_x g_i(\tilde{x}_n, u_0) h_n \leqslant o(1)$, 因而有 $D_x g_i(x_0, u_0) h \leqslant 0$. 类似可得 $D_x g_i(x_0) h = 0$ 对所有 $i \leqslant q$ 成立. 可见 h 是临界方向, 即 $h \in T_{\Theta}(x_0)$. 不等式 (4.382) 由 (4.381) 可得到, 这导致矛盾. □

注意到, 由于 $C(x_0) = T_{\Theta}(x_0)$, 标准的二阶必要条件可以表示为

$$D_{xx}^2 L(x_0, \lambda(x_0), u_0)(h, h) \geqslant 0, \quad \forall\, h \in T_{\Theta}(x_0). \tag{4.384}$$

若 x_0 是 $\mathcal{S}(u_0)$ 的孤立点, 则 $T_{\mathcal{S}}(x_0) = \{0\}$, 此种情况, 二阶条件 (4.381) 取标准形式, 即可以通过将二阶条件 (4.384) 中的 "$\geqslant$" 替换为严格不等式符号 "$>$" 而得到.

问题 P_u 在点 (x_0, u_0) 处的线性化可表示为下述形式

$$(\mathrm{PL}_d(x_0)) \qquad \begin{aligned} \min_{h \in X} \quad & Df(x_0, u_0)(h, d), \\ \text{s. t.} \quad & Dg_i(x_0, u_0)(h, d) = 0, \quad i = 1, \cdots, q, \\ & Dg_i(x_0, u_0) \leqslant 0, \quad i \in I(x_0, u_0). \end{aligned}$$

在线性无关的约束规范下, 上述问题 $(\mathrm{PL}_d(x_0))$ 的可行点 h 是最优解当且仅当

$$\lambda_i(x_0) Dg_i(x_0, u_0)(h, d) = 0, \quad i \in I(x_0, u_0), \tag{4.385}$$

---

① 下式应为 $f(x_n, u_0) \leqslant f(x_0, u_0) + o\Big(\operatorname{dist}(x_n, \mathcal{S}(u_0))^2\Big)$.

② 原著中为 $\mathcal{S}(x_0)$.

③ 原著中为 $\mathcal{S}(x_0)$.

④ 原著中为 $\mathcal{S}(x_0)$.

由对偶性有

$$\mathrm{val}(\mathrm{PL}_d(x_0)) = D_u L(x_0, \lambda(x_0), u_0)d. \tag{4.386}$$

若还有严格互补条件成立, 则 $(\mathrm{PL}_d(x_0))$ 的最优解集变为

$$\mathcal{S}(\mathrm{PL}_d(x_0)) = \{h : Dg_i(x_0, u_0)(h, d) = 0, i \in \{1, \cdots, q\} \cup I(x_0, u_0)\}. \tag{4.387}$$

此种情形, 集合 $\mathcal{S}(\mathrm{PL}_d(x_0))$ 是平行于切空间 $T_\Theta(x_0)$ 的仿射空间.

定理 4.142 设

(i) 线性无关的约束规范在每一点 $x_0 \in \mathcal{S}(u_0)$ 处均成立.

(ii) 下确界–紧致性条件成立.

(iii) 严格互补条件在每一点 $x_0 \in \mathcal{S}_1(u_0, d)$ 处均成立.

(iv) 在每一 $x_0 \in \mathcal{S}_1(u_0, d)$ 的邻域中, $\mathcal{S}(u_0)$ 均是 $\Theta(x_0, u_0)$ 的光滑子流形.

(v) 在每一 $x_0 \in \mathcal{S}_1(u_0, d)$ 处二阶条件 (4.381) 成立.

则下述结论成立:

(a) 对任何形式为 $u(t) := u_0 + td + o(t)$ 的路径, 当 $t \downarrow 0$ 时, 多值函数 $t \to \mathcal{S}(u(t))$ 是上 Lipschitz 连续的.

(b) 最优值函数 $v(u)$ 沿方向 d 是二阶方向可微的, 且

$$v''(u_0; d, r) = \inf_{x \in \mathcal{S}_1(u_0, d)} \{D_u L(x, \lambda(x), u_0)r + \xi_d(x)\}, \tag{4.388}$$

其中

$$\xi_d(x) := \inf_{h \in \mathcal{S}(\mathrm{PL}_d(x))} D^2 L(x, \lambda(x), u_0)((h, d), (h, d)). \tag{4.389}$$

(c) 当 $t \downarrow 0$ 时, 对任何满足 $tr(t) \downarrow 0$ 的路径 $r(\cdot) : \mathbb{R}_+ \to U$, 下述不等式成立

$$v\left(u_0 + td + \frac{1}{2}t^2 r(t)\right) \leqslant v(u_0) + tv'(u_0, d) + \frac{t^2}{2} v''(u_0; d, r(t)) + o(t^2). \tag{4.390}$$

(d) 对任何路径 $u(t) := u_0 + td + \frac{t^2}{2} r + o(t^2)$ 与 $(\mathrm{P}_{u(t)})$ 的最优解 $\bar{x}(t)$, 当 $t \downarrow 0$ 时, $\bar{x}(t)$ 的任何极限点均包含在集合

$$\mathcal{S}_2(u_0; d, r) := \operatorname*{argmin}_{x \in \mathcal{S}_1(u_0, d)} \{D_u L(x, \lambda(x), u_0)r + \xi_d(x)\} \tag{4.391}$$

中, 进一步, 若 $\tilde{x}(t)$ 是 $\mathcal{S}(u_0)$ 中与 $\bar{x}(t)$ 最近的点, 则存在 $x \in \mathcal{S}_2(u_0; d, r)$, 满足 $t^{-1}(\bar{x}(t) - \tilde{x}(t))$ 的每一极限点均是 (4.389) 右端的最优解.

证明 考虑路径 $u(t) := u_0 + td + o(t)$, $t \geqslant 0$, 令 $\bar{x}(t)$ 是 $(\mathrm{P}_{u(t)})$ 的最优解. 由命题 4.140, 当 $t \downarrow 0$ 时, $\bar{x}(t)$ 的每一极限点均属于集合 $\mathcal{S}_1(u_0, d)$. 根据命题 4.141, 由假设

可得, 在 $\mathcal{S}_1(u_0,d)$ 的每一点的邻域内二阶增长条件成立. 由于线性无关的约束规范可推出非退化性 (见定义 4.70 与例 3.139), 由定理 4.81 可得 $\operatorname{dist}(\bar{x}(t),\mathcal{S}(u_0))=O(t)$, 这证得 (a).

由命题 4.140 与 (4.386), 对任何 $x\in\mathcal{S}_1(u_0,d)$, $v'(u_0,d)$ 等于问题 $(\mathrm{PL}_d(x))$ 的最优值. 由定理 4.85 的上方估计 (4.217) 可得上二阶方向导数 $v''_+(u_0;d,r)$ 小于或等于 (4.388) 的右端.

现在考虑一路径 $u(t):=u_0+td+\dfrac{t^2}{2}r+o(t^2)$ 与一序列 $t_n\downarrow 0$, 置 $u_n:=u(t_n)$. 令 $x_n\in\mathcal{S}(u_n)$, $\tilde{x}_n$ 是 $\mathcal{S}(u_0)$ 中距离 x_n 最近的点, 置 $h_n:=t_n^{-1}(x_n-\tilde{x}_n)$. 由下确界–紧致性条件, 序列 $\{x_n\}$ 是有界的, 由结论 (a) 可得序列 $\{h_n\}$ 是有界的. 因此, 若有必要取一子序列, 不妨设 $\{x_n\}$ 与 $\{h_n\}$ 分别收敛到某点 x_0 与 h. 由命题 4.140 可得 $x_0\in\mathcal{S}_1(u_0,d)$. 进一步, 由严格互补条件, 对充分接近于 u_0 的 u 与充分接近于 x_0 的问题 (P_u) 的最优解 $\bar{x}(u)$, 对应于 $i\in I(x_0,u_0)$ 的不等式约束在 $\bar{x}(u)$ 处是起作用的. 因此, 对充分大的 n 有 $g_i(x_n,u_n)=0$ 与 $g_i(\tilde{x}_n,u_0)=0$, 从而对所有 $i\in\{1,\cdots,q\}\cup I(x_0,u_0)$ 有 $Dg_i(x_0,u_0)(h,d)=0$. 进一步, 对充分大的 n, $v(u_0)=L(\tilde{x}_n,\lambda(\tilde{x}_n),u_0)$, $v(u_n)=L(x_n,\lambda(\tilde{x}_n),u_n)$. 结果, 由于 $D_xL(\tilde{x}_n,\lambda(\tilde{x}_n),u_0)=0$, 有

$$\begin{aligned}v(u_n)-v(u_0)&=L(x_n,\lambda(\tilde{x}_n),u_n)-L(\tilde{x}_n,\lambda(\tilde{x}_n),u_0)\\&=t_nD_uL(\tilde{x}_n,\lambda(\tilde{x}_n),u_0)d+\frac{1}{2}t_n^2D_uL(\tilde{x}_n,\lambda(\tilde{x}_n),u_0)r\\&\quad+\frac{1}{2}t_n^2D^2L(\tilde{x}_n,\lambda(\tilde{x}_n),u_0)((h_n,d),(h_n,d))+o(t_n^2).\end{aligned}\tag{4.392}$$

由 (4.378) 有

$$v'(u_0,d)\leqslant D_uL(\tilde{x}_n,\lambda(\tilde{x}_n),u_0)d.\tag{4.393}$$

因为 $h\in\mathcal{S}(\mathrm{PL}_d(x_0))$, 由 (4.392) 与 (4.393) 可得

$$v(u_n)-v(u_0)-t_nv'(u_0,d)\geqslant\frac{1}{2}t_n^2[D_uL(x_0,\lambda(x_0),u_0)r+\xi_d(x_0)]+o(t_n^2).\tag{4.394}$$

得到下二阶方向导数 $v''_-(u_0,d,r)$ 大于或等于 (4.388) 的右端. 结果, 二阶方向导数 $v''(u_0,d,r)$ 存在, 且公式 (4.388) 成立. 这证得 (b). 由 $x_0\in\mathcal{S}_2(u_0;d,r)$ 及 h 是 $D^2L(x_0,\lambda(x_0),u_0)((\cdot,d),(\cdot,d))$ 在 $\mathcal{S}(\mathrm{PL}_d(x_0))$ 上的极小点. 证得 (d).

证明 (c). 考虑当 $t\downarrow 0$ 时, 路径 $u(t):=u_0+td+\dfrac{t^2}{2}r(t)$ 满足 $tr(t)\to 0$, $x_0\in\mathcal{S}_1(u_0,d)$, $h\in\mathcal{S}(\mathrm{PL}_d(x_0))$. 由线性无关的约束规范, 根据隐函数定理, 对 $t\geqslant 0$ 充分小, 存在 $x(t)$ 满足 $g_i(x(t),u(t))=0$, $i\in\{1,\cdots,q\}\cup I(x_0,u_0)$, $x(t)=x_0+th+\varepsilon(t)$, 其中 $\varepsilon(t)=o(t)$ 关于 $x_0\in\mathcal{S}_1(u_0,d)$ 是一致的. 注意到, 由连续性条件, 对于 $i\in\{q+1,\cdots,p\}\setminus I(x_0,u_0)$, $g_i(x(t),u(t))<0$, 因此对 $t\geqslant 0$ 充分小, $x(t)$ 是 $(\mathrm{P}_{u(t)})$

的可行点. 从而对充分小的 $t \geqslant 0$, 有

$$
\begin{aligned}
v(u(t)) - v(u_0) \leqslant & L(x(t), \lambda(x_0), u(t)) - L(x_0, \lambda(x_0), u_0) \\
= & tD_u L(x_0, \lambda(x_0), u_0) + \frac{1}{2}t^2 D_u L(x_0, \lambda(x_0), u_0) r(t) \\
& + \frac{1}{2}t^2 D^2 L(x_0, \lambda(x_0), u_0)((h,d),(h,d)) + \gamma(t),
\end{aligned} \tag{4.395}
$$

其中余项 $\gamma(t)$ 是 $o(t^2)$ 阶的, 关于 $x_0 \in \mathcal{S}_1(u_0, d)$ 是一致的. 最后, 由 (b) 与 (4.392) 可得到 (d). □

注 4.143 不等式 (4.390) 意味着在上述定理的假设下, $-v(\cdot)$ 在 u_0 点处沿方向 d 是二阶上图正则的 (见定义 3.94).

注 4.144 由二阶必要条件 (4.384), 则二次型

$$
\kappa(\cdot) := D^2 L(x_0, \lambda(x_0), u_0)((\cdot, d), (\cdot, d))
$$

在线性空间 $T_\Theta(x_0)$ 上是半正定的. 然而在 x_0 的邻域内, $L(\cdot, \lambda(x_0), u_0)$ 在 $\mathcal{S}(u_0)$ 上取常值, 从而

$$
D^2_{xx} L(x_0, \lambda(x_0), u_0)(h,h) = 0, \quad \forall h \in T_{\mathcal{S}}(x_0). \tag{4.396}
$$

所以, $\kappa(\cdot)$ 在 $T_\Theta(x_0)$ 上不是正定的, 除非 $T_{\mathcal{S}}(x_0) = \{0\}$, 即 x_0 是 $\mathcal{S}(u_0)$ 的孤立点的情形. 结果可能发生这样的情况, 存在某一 $x \in \mathcal{S}(u_0)$, 使得 (4.389) 中定义的极小值函数 $\xi_d(x)$ 取 $-\infty$ 值. 然而, 对任何 $x_0 \in \mathcal{S}_1(u_0, d)$, 由相应的最优化问题的一阶必要性条件可得

$$
D^2_{xu} L(x_0, \lambda(x_0), u_0)(h,d) = 0, \quad \forall h \in T_{\mathcal{S}}(x_0). \tag{4.397}
$$

结合二阶条件 (4.381) 可推出 $\kappa(h)$ 在 $T_\Theta(x_0)$ 上取得其极小值, 从而在平行的空间 $\mathcal{S}(\mathrm{PL}_d(x_0))$ 上取得极小值. 所以在定理 4.142 的假设下, $\xi_d(x_0)$ 是有限的, (4.389) 右端的极小是可取到的, $v''(u_0; d, r)$ 是有限值的, 集合 $\mathcal{S}_2(u_0; d, r)$ 是非空的. 这也可以从定理 4.142 中的结论 (a) 与 (d) 非直接地得到. 最后注意到, 由 (4.396) 与 (4.397), $\kappa(\cdot)$ 在平行于 $T_{\mathcal{S}}(x_0)$ 的任何空间上都取常值, 所以 $\kappa(\cdot)$ 在 $\mathcal{S}(\mathrm{PL}_d(x_0))$ 上的极小点集合是平行于 $T_{\mathcal{S}}(x_0)$ 的 $\mathcal{S}(\mathrm{PL}_d(x_0))$ 的子空间, 因此除非 $T_{\mathcal{S}}(x_0) = \{0\}$, 它不会是单点集.

例 4.145 设 $U := \mathcal{S}^p$ 是 $p \times p$ 对称矩阵空间. 则对于对称矩阵 $A \in \mathcal{S}^p$, 其最大特征值 $\lambda_{\max}(A)$ 可以表示为下述形式

$$
\lambda_{\max}(A) = \max_{\|x\|=1} x^{\mathrm{T}} A x. \tag{4.398}
$$

很清楚, 函数 $f(x,A) := x^{\mathrm{T}}Ax$ 是二次连续可微的, 可行集 $\Phi := \{x \in \mathbb{R}^p : \|x\| = 1\}$ 是球, 因而是紧致的光滑流形. (4.398) 中的最大值在对应于最大特征值 $\lambda_{\max}(A)$ 的 A 的任何长度是 1 个单位的特征向量处达到. 令 s 是 $\lambda_{\max}(A)$ 的重数, E 是 $p \times s$ 矩阵, 它的列 $e_1, \cdots, e_s$ 由对应于最大特征值的 A 的特征向量空间的一组直交基构成. 问题 (4.398) 的最优解集由下式给出:

$$\mathcal{S}(A) = \{x = E\beta : \beta \in \mathbb{R}^s, \|\beta\| = 1\}. \tag{4.399}$$

由例 4.20

$$\lambda'_{\max}(A,H) = \lambda_{\max}(E^{\mathrm{T}}HE).$$

令 r 是 $E^{\mathrm{T}}HE$ 的最大特征值的重数, F 是 $s \times r$ 矩阵, 它的列 $f_1, \cdots, f_r$ 由对应最大特征值的 $E^{\mathrm{T}}HE$ 的特征向量空间的一组直交基构成. 则

$$\mathcal{S}_1(A,H) = \{x = EF\gamma : \gamma \in \mathbb{R}^r, \|\gamma\| = 1\}. \tag{4.400}$$

进一步

$$L(x,\alpha,A) := x^{\mathrm{T}}Ax - \alpha x^{\mathrm{T}}x,$$

对所有的 $x \in \mathcal{S}(A)$, 相应的 Lagrange 乘子 $\alpha(x)$ 等于 $\lambda_{\max}(A)$. 由于 $L(x,\alpha,A)$ 关于 A 是线性的, 且

$$\begin{aligned} D^2L(x,\alpha,A)((h,H),(h,H)) &= 2h^{\mathrm{T}}Ah + 4x^{\mathrm{T}}Hh - 2\alpha h^{\mathrm{T}}h \\ &= 2h^{\mathrm{T}}(A - \alpha I_p)h + 4x^{\mathrm{T}}Hh, \end{aligned}$$

$\mathcal{S}(\mathrm{PL}_H(x)) = \{h : h^{\mathrm{T}}x = 0\}$. 由公式 (4.388) 得到 (注意, 这里的最优值函数是最大值函数而不是最小值函数)

$$\lambda''_{\max}(A; H, W) = \sup_{x \in \mathcal{S}_1(A,H)} \{x^{\mathrm{T}}Wx + \xi_H(x)\}, \tag{4.401}$$

其中 $\bar{\alpha} := \lambda_{\max}(A)$,

$$\xi_H(x) = \sup_{h^{\mathrm{T}}x=0} \{2h^{\mathrm{T}}(A - \bar{\alpha}I_p)h + 4x^{\mathrm{T}}Hh\}.$$

令 $(A - \bar{\alpha}I_p)^{\dagger}$ 记矩阵 $A - \bar{\alpha}I_p$ 的 Moore-Penrose 伪逆, 即

$$(A - \bar{\alpha}I_p)^{\dagger} = \sum_{\lambda_i \neq \bar{\alpha}} (\lambda_i - \bar{\alpha})^{-1} a_i a_i^{\mathrm{T}},$$

其中 λ_i 是 A 的特征值, a_i 是相应的正交的特征向量. 因为对 $x \in \mathcal{S}_1(A,H)$, $x^{\mathrm{T}}x = 1$ 且 $(A - \bar{\alpha}I_p)x = 0$, 且由于

$$(A - \bar{\alpha}I_p)(A - \bar{\alpha}I_n)^{\dagger}(A - \bar{\alpha}I_p) = (A - \bar{\alpha}I_p),$$

则对 $x \in \mathcal{S}_1(A, H)$,

$$\xi_H(x) = -2x^{\mathrm{T}} H(A - \bar{\alpha} I_p)^{\dagger} Hx. \tag{4.402}$$

结果有

$$\lambda''_{\max}(A; H, W) = \sup_{x \in \mathcal{S}_1(A,H)} \{x^{\mathrm{T}}[W - 2H(A - \bar{\alpha} I_p)^{\dagger} H]x\}.$$

由 (4.400) 可以推出

$$\lambda''_{\max}(A; H, W) = \lambda_{\max}(F^{\mathrm{T}} E^{\mathrm{T}}[W - 2H(A - \bar{\alpha} I_p)^{\dagger} H]EF). \tag{4.403}$$

进一步, 由定理 4.142(c), 对任何满足当 $t \downarrow 0$, $tW(t) \to 0$ 的 $W(\cdot) : \mathbb{R}_+ \to \mathcal{S}^p$, 不等式

$$\begin{aligned}\lambda_{\max}\left(A + tH + \frac{1}{2}t^2 W(t)\right) \geqslant\ & \lambda_{\max}(A) + t\lambda'_{\max}(A, H) \\ & + \frac{1}{2}t^2 \lambda''_{\max}(A; H, W(t)) + o(t^2)\end{aligned}$$

对 $t \geqslant 0$ 成立. 即函数 $\lambda_{\max}(\cdot)$ 是二阶上图正则的, 因此是二阶正则的 (见定义 3.93 与命题 3.95). 于是空间 $\mathcal{S}^p$ 中正 (负) 半定矩阵构成的锥也是二阶正则的 (见命题 3.92 与 (3.169)). 在例 3.140 中已经用简约化方法证得这一结论.

4.10 泛函空间中的二阶分析

这一节在泛函空间讨论二阶切集与最优值函数二阶展开的计算. 我们将看到, 二阶切集的公式与最优值函数的二阶方向导数公式是紧密联系的.

4.10.1 连续函数的泛函空间的二阶切集

设 Ω 是紧致空间, 令 $C(\Omega)$ 是定义于 Ω 上的, 赋予 sup 范数的连续函数的 Banach 空间. 这一节讨论下述非负值函数构成的锥的二阶切集

$$C_+(\Omega) := \{x \in C(\Omega) : x(\omega) \geqslant 0, \forall\, \omega \in \Omega\}.$$

由例 2.63, 在点 $x \in C_+(\Omega)$ 处的锥 $K := C_+(\Omega)$ 的切锥为

$$T_K(x) = \{z \in C(\Omega) : z(\omega) \geqslant 0, \forall\, \omega \in \Delta(x)\}, \tag{4.404}$$

其中

$$\Delta(x) := \{\omega \in \Omega :\ x(\omega) = 0\} \tag{4.405}$$

是 x 处的接触点集.

为继续讨论, 需要下述引理. 用 Σ 记收敛到零的正实数序列 $\sigma = \{t_n\}$ 的集合. 若 $\{f_t\}$ 是 $C(\Omega)$ 中的一族函数, 则当 $t \to t_0$ 时 $\{f_t\}$ 下上图极限可表示为

$$\mathrm{e} - \liminf_{t \to t_0} f_t(\omega) = \liminf_{\substack{t \to t_0 \\ \omega' \to \omega}} f_t(\omega'). \tag{4.406}$$

引理 4.146　*设 Ω 是紧致度量空间, 令 $\{f_n\}$ 是 $C(\Omega)$ 中的序列. 则*

$$\liminf_{n \to \infty} (\inf_{\omega \in \Omega} f_n(\omega)) = \inf_{\omega \in \Omega} (\mathrm{e} - \liminf_{n \to \infty} f_n(\omega)). \tag{4.407}$$

证明　分别用 α 与 β 记 (4.407) 中左端的值与右端的值, 则存在子序列 $n_k \in N$ 与 $\omega_k \in \Omega$ 满足 $\alpha = \lim_{n \to \infty} f_{n_k}(\omega_k)$. 因为序列 $\{\omega_k\}$ 有极限点 $\bar{\omega} \in \Omega$, 则 $\beta \leqslant \mathrm{e} - \lim_{n\to\infty} \inf f_n(\bar{\omega}) \leqslant \alpha$. 另一方面, 对任意 $\varepsilon > 0$ 与 $\omega \in \Omega$ 及充分大的 n, 有 $f_n(\omega) \geqslant \alpha - \varepsilon$, 因此有 $\mathrm{e} - \lim_{n \to \infty} \inf f_n(\omega) \geqslant \alpha - \varepsilon$. 这意味着 $\beta \geqslant \alpha - \varepsilon$. 由于这对任何 $\varepsilon > 0$ 均是成立的, 有 $\beta \geqslant \alpha$, 结论得证. □

定理 4.147　*令 Ω 是紧致的度量空间, $K := C_+(\Omega)$, $x \in K$, $h \in T_K(x)$, $\sigma := \{t_n\} \in \Sigma$, 考虑如下定义的增广实值函数 $\tau^\sigma_{x,h}(\cdot)$:*

$$\tau^\sigma_{x,h} := -\mathrm{e} - \liminf_{n \to \infty} \left\{ \frac{x + t_n h}{t_n^2/2} \right\}. \tag{4.408}$$

则

$$T_K^{i,2,\sigma}(x,h) = \{z \in C(\Omega) : z(\omega) \geqslant \tau^\sigma_{x,h}(\omega),\ \forall \omega \in \Omega\}, \tag{4.409}$$

且 $T_K^{i,2,\sigma}(x,h)$ 非空的充分必要条件是对所有的 $\omega \in \Omega$, $\tau^\sigma_{x,h}(\omega) < +\infty$.

证明　注意到下上图极限函数是下半连续的, 由 (4.408) 定义的函数 $\tau^\sigma_{x,h}(\cdot)$ 是上半连续的. 由定义有 $z \in T_K^{i,2,\sigma}(x,h)$ 当且仅当 $\mathrm{dist}\left(x + t_n h + \frac{1}{2} t_n^2 z, K\right) = o(t_n^2)$. 对任何 $y \in C(\Omega)$, 置 $a_+ := \max\{a, 0\}$, 有

$$\mathrm{dist}(y, C_+(\Omega)) = \sup_{\omega \in \Omega} \{[-y(\omega)]_+\}.$$

令

$$f_n(\omega) := z(\omega) + \frac{x(\omega) + t_n h(\omega)}{t_n^2/2}.$$

则 $z \in T_K^{i,2,\sigma}(x,h)$ 当且仅当 $\inf_{\omega \in \Omega} f_n(\omega)) \geqslant o(1)$, 或等价地,

$$\liminf_{n \to \infty} (\inf_{\omega \in \Omega} f_n(\omega)) \geqslant 0.$$

将此式与引理 4.146 相结合并由于

$$\mathrm{e}-\liminf_{n\to\infty} f_n = z + e - \liminf_{n\to\infty}\left\{\frac{x+t_n h}{t_n^2/2}\right\},$$

即得到 (4.409).

若存在 $\omega\in\Omega$, 有 $\tau^{\sigma}_{x,h}(\omega)=+\infty$, 则由 (4.409) 可得 $T^{i,2,\sigma}_K(x,h)$ 是空集. 相反地, 由于 $\tau^{\sigma}_{x,h}(\cdot)$ 是上半连续的且 Ω 是紧致的, 若对所有的 $\omega\in\Omega$, 有 $\tau^{\sigma}_{x,h}(\omega)<+\infty$, 则 $\tau^{\sigma}_{x,h}(\cdot)$ 是上方一致有界的, 因此, 由 (4.409) 得 $T^{i,2,\sigma}_K(x,h)$ 是非空的. □

因内二阶切集 $T^{i,2}_K(x,h)$ 由集合 $T^{i,2,\sigma}_K(x,h)$ 取所有序列 $\sigma\in\Sigma$ 的交得到, 外二阶切集$T^2_K(x,h)$ 是集合 $T^{i,2,\sigma}_K(x,h)$ 取所有序列 $\sigma\in\Sigma$ 的并得到. 于是得到这些二阶切集的下述刻画.

定理 4.148 设 Ω 是紧致度量空间, $K:=C_+(\Omega)$, $x\in K$, $h\in T_K(x)$. 则

(i) 内二阶切集可表示为下述形式

$$T^{i,2}_K(x,h)=\{z\in C(\Omega): z(\omega)\geqslant \tau_{x,h}(\omega),\ \forall\,\omega\in\Omega\}, \tag{4.410}$$

其中函数 $\tau_{x,h}:\Omega\to\overline{\mathbb{R}}$ 定义为

$$\tau_{x,h}:=-\mathrm{e}-\liminf_{t\downarrow 0}\left\{\frac{x+th}{t^2/2}\right\}, \tag{4.411}$$

且 $T^{i,2}_K(x,h)$ 非空的充分必要条件是对所有的 $\omega\in\Omega$ 有 $\tau_{x,h}(\omega)<+\infty$.

(ii) 若 $z\in T^2_K(x,h)$, 则 $z(\omega)\geqslant \tau^*_{x,h}(\omega)$, $\forall\,\omega\in\Omega$, 其中

$$\tau^*_{x,h}:=-\mathrm{e}-\limsup_{t\downarrow 0}\left\{\frac{x+th}{t^2/2}\right\}. \tag{4.412}$$

(iii) $T^{i,2}_K(x,h)=T^2_K(x,h)$ 成立当且仅当函数 $f_t:=\dfrac{x+th}{t^2/2}$ 当 $t\downarrow 0$ 时是上图收敛的.

证明 结论 (i) 用类似于定理 4.147 的证明方法证明. 我们证 (ii). $z\in T^2_K(x,h)$ 当且仅当存在 $\sigma\in\Sigma$ 满足 $z\in T^{i,2,\sigma}_K(x,h)$, 由定理 4.147, 这等价于 $z(\cdot)\geqslant\tau^{\sigma}_{x,h}(\cdot)$. 于是, 若 $z\in T^2_K(x,h)$, 则

$$z\geqslant\inf_{\sigma\in\Sigma}\tau^{\sigma}_{x,h}=-\sup_{\sigma\in\Sigma}\left(\mathrm{e}-\liminf_{n\to+\infty}\left\{\frac{x+t_nh}{t_n^2/2}\right\}\right)=-\mathrm{e}-\limsup_{t\downarrow 0}\left\{\frac{x+th}{t^2/2}\right\}$$ [①]

这证得 (ii).

$T^{i,2}_K(x,h)=T^2_K(x,h)$ 当且仅当对任何 $\sigma\in\Sigma$ 有 $T^{i,2,\sigma}_K(x,h)=T^2_K(x,h)$. 由 (4.409) 与 (4.410), 这一等式成立的充要条件是对所有的 $\sigma\in\Sigma$ 有 $\tau^{\sigma}_{x,h}(\cdot)=\tau_{x,h}(\cdot)$, 这等价于 $\tau^*_{x,h}(\cdot)=\tau_{x,h}(\cdot)$. 这证得 (iii). □

①最后一式原著为 $n\to+\infty$, 应该为 $t\downarrow 0$.

现在给出定理 4.147 中用到的函数 $\tau_{x,h}(\cdot)$ 的更加显式的公式. 集合 $\Delta\subset\Omega$ 的边界定义为

$$\mathrm{bdr}(\Delta):=\mathrm{cl}(\Delta)\backslash\mathrm{int}(\Delta). \tag{4.413}$$

尤其, 若集合 Δ 是闭的, 则 $\mathrm{bdr}(\Delta)=\Delta\backslash\mathrm{int}(\Delta)$.

定理 4.149 *设 Ω 是紧致的度量空间, $K:=C_+(\Omega)$, $x\in K$, $h\in T_K(x)$. 则由 (4.411) 定义的函数 $\tau_{x,h}(\omega)$ 可以表示为下述形式*

$$\tau_{x,h}(\omega)=\begin{cases}0, & \text{若 } \omega\in\mathrm{int}(\Delta(x)) \text{ 且 } h(\omega)=0,\\ \theta(\omega), & \text{若 } \omega\in\mathrm{bdr}(\Delta(x)) \text{ 且 } h(\omega)=0,\\ -\infty, & \text{否则},\end{cases} \tag{4.414}$$

其中

$$\theta(\omega):=\limsup_{\substack{\omega'\to\omega\\ x(\omega')>0}}\frac{([-h(\omega')]_+)^2}{2x(\omega')}. \tag{4.415}$$

证明 注意到, x 的接触点集 $\Delta(x)$ 是闭的, 因此它是两个互不相交的集合 $\mathrm{int}(\Delta(x))$ 与 $\mathrm{bdr}(\Delta(x))$ 的并集. 先考虑 $x(\omega)>0$ 的情况. 则存在数 $\varepsilon>0$ 与 δ 满足对充分接近于 ω 的所有的 ω', 有 $x(\omega')\geqslant\varepsilon$ 与 $h(\omega')\geqslant\delta$. 于是

$$\mathrm{e}-\liminf_{t\downarrow 0}\frac{x(\omega)+th(\omega)}{t^2/2}=\liminf_{\substack{t\downarrow 0\\ \omega'\to\omega}}\frac{x(\omega')+th(\omega')}{t^2/2}\geqslant\liminf_{t\downarrow 0}\frac{\varepsilon+t\delta}{t^2/2}=+\infty,$$

因此有 $\tau_{x,h}(\omega)=-\infty$. 类似地, 若 $x(\omega)=0$ 且 $h(\omega)>0$, 在前述不等式中可取 $\varepsilon=0$, $\delta>0$, 从而得到 $\tau_{x,h}(\omega)=-\infty$. 可见 "否则" 情形的表达式成立.

现在设 $\omega\in\mathrm{int}(\Delta(x))$, $h(\omega)=0$. 由于 $h\in T_K(x)$, 对所有的 $\omega\in\Delta(x)$, 有 $h(\omega)\geqslant 0$, 于是对充分接近于 ω 的 ω', 有 $x(\omega')=0\leqslant h(\omega')$, 从而有

$$\liminf_{\substack{t\downarrow 0\\ \omega'\to\omega}}\frac{x(\omega')+th(\omega')}{t^2/2}=\liminf_{\substack{t\downarrow 0\\ \omega'\to\omega}}\frac{2h(\omega')}{t}=0.$$

这就得到 $\tau_{x,h}(\omega)=0$.

最后考虑 $\omega\in\mathrm{bdr}(\Delta(x))$ 且 $h(\omega)=0$ 的情况. 因为集合 $\Delta(x)$ 是闭的, 有 $\mathrm{bdr}(\Delta(x))\subset\Delta(x)$, 则 $x(\omega)=0$, 因为 $\omega\notin\mathrm{int}(\Delta(x))$, 存在与 ω 任意接近的 ω' 满足 $x(\omega')>0$. 考虑函数

$$\phi(t):=\frac{x(\omega')+th(\omega')}{t^2/2},\quad t>0.$$

若 $h(\omega')\geqslant 0$, 则 $\phi(t)>0$, 且当 $t\to+\infty$ 时, $\phi(t)\to 0$, 因此有 $\inf_{t>0}\phi(t)=0$. 若 $h(\omega')<0$, 则

$$\inf_{t>0}\phi(t)=\frac{-h(\omega')^2}{2x(\omega')},$$

其最小值在 $t = -2x(\omega')/h(\omega')$ 处达到. 所以,

$$-\theta(\omega) = -\limsup_{\substack{\omega' \to \omega \\ x(\omega')>0}} \frac{([-h(\omega')]_+)^2}{2x(\omega')} \leqslant \liminf_{\substack{t \downarrow 0 \\ \omega' \to \omega}} \frac{x(\omega') + th(\omega')}{t^2/2},$$

因此有 $\theta(\omega) \geqslant \tau_{x,h}(\omega)$.

下面证明相反的不等式成立. 注意 $\tau_{x,h}(\omega) \geqslant 0$, 在 $\tau_{x,h}$ 的定义中, 估计下导数时取 $\omega' = \omega$ 可得到这一不等式. 若 $\theta(\omega) = 0$, 则可得 $\theta(\omega) \leqslant \tau_{x,h}(\omega)$. 因此, 可设 $\theta(\omega) > 0$. 令 $\omega_n \to \omega$ 是 $\theta(\omega)$ 的定义中实现其上极限的序列. 因为 $\theta(\omega) > 0$, 对充分大的 n, 有 $h(\omega_n) < 0$. 置 $t_n := -2x(\omega_n)/h(\omega_n)$. 则 $-h(\omega_n)/t_n \to \theta(\omega) > 0$, 因此 $t_n \downarrow 0$. 于是

$$\begin{aligned}\liminf_{\substack{t \downarrow 0 \\ \omega' \to \omega}} \frac{x(\omega') + th(\omega')}{t^2/2} &\leqslant \liminf_{n \to \infty} \frac{x(\omega_n) + t_n h(\omega_n)}{t_n^2/2} \\ &= \lim_{n \to \infty} \frac{-h(\omega_n)^2}{2x(\omega_n)} = -\theta(\omega),\end{aligned}$$

因此 $\theta(\omega) \leqslant \tau_{x,h}(\omega)$. 结果有 $\theta(\omega) = \tau_{x,h}(\omega)$, 证毕. □

注 4.150 设 $h \in T_K(x)$, 即对所有的 $\omega \in \Delta(x)$, $h(\omega) \geqslant 0$, $h(\cdot)$ 在 Ω 上是 Lipschitz 连续的, 下述二阶增长条件成立：存在 $\alpha > 0$, 满足

$$x(\omega) \geqslant \alpha[\operatorname{dist}(\omega, \Delta(x))]^2, \quad \forall \omega \in \Omega. \tag{4.416}$$

对满足 $h(\omega) = 0$ 的任何 $\omega \in \operatorname{bdr}(\Delta(x))$, $\theta(\omega)$ 是有限的, 因此 $\tau_{x,h}(\cdot) < +\infty$, 由定理 4.147, 内二阶切集 $T_K^{i,2}(x,h)$ 非空.

现在给出使内二阶切集与外二阶切集相等的充分条件. 因 Ω 的子集 Γ 是连通的, 若 Γ 中的任何两点均可以由连续的曲线连接, 即对任意 $\omega_1, \omega_2 \in \Gamma$, 存在连续映射 $\phi: [0,1] \to \Gamma$ 满足 $\phi(0) = \omega_1$ 与 $\phi(1) = \omega_2$.

命题 4.151 *设 Ω 是紧致的度量空间, $K := C_+(\Omega)$, $x \in K$, $h \in T_K(x)$, 设对任何满足 $h(\omega) = 0$ 与 $\theta(\omega) > 0$ 的 $\omega \in \operatorname{bdr}(\Delta(x))$, 存在连通的集合 $\Gamma_\omega \subset \Omega$, 满足 ω 是 $\Gamma_\omega \subset \Omega$ 的极限点, 对任意 $\omega' \in \Gamma_\omega$, 有 $x(\omega') > 0$, 且*

$$\limsup_{\substack{\omega' \to \omega \\ x(\omega')>0}} \frac{([-h(\omega')]_+)^2}{2x(\omega')} = \liminf_{\substack{\omega' \to \omega \\ \omega' \in \Gamma_\omega}} \frac{([-h(\omega')]_+)^2}{2x(\omega')}. \tag{4.417}$$

则 $\tau_{x,h}(\cdot) = \tau^*_{x,h}(\cdot)$, 因此有 $T_K^{i,2}(x,h) = T_K^2(x,h)$.

证明 用类似于定理 4.149 的推证易得, 若 $\omega \in \operatorname{int}(\Delta(x))$ 且 $h(\omega) = 0$, 则 $\tau^*_{x,h}(\omega) = 0$, 若 $x(\omega) > 0$ 或者 $x(\omega) = 0$ 且 $h(\omega) > 0$, 则 $\tau^*_{x,h}(\omega) = -\infty$. 因此, 除了可能的 $\omega \in \operatorname{bdr}(\Delta(x))$ 且 $h(\omega) = 0$ 外, 均有 $\tau^*_{x,h}(\omega) = \tau_{x,h}(\omega)$.

令 $\omega \in \operatorname{bdr}(\Delta(x))$ 满足 $h(\omega)=0$. 考虑序列 $\sigma \in \Sigma$. 在任何情形均有 $\theta(\omega)=\tau_{x,h}(\omega) \geqslant \tau_{x,h}^{\sigma}(\omega) \geqslant 0$, 因此, 若 $\theta(\omega)=0$, 则 $\tau_{x,h}(\omega)=\tau_{x,h}^{\sigma}(\omega)=0$. 所以, 可设 $\theta(\omega)>0$. 令 $\bar{\omega}_n \to \omega$ 是 (4.417) 的左端的上极限被实现的序列. 注意到条件 (4.417) 意味着极限

$$\lim_{\substack{\omega' \to \omega \\ \omega' \in \Gamma_\omega}} \frac{([-h(\omega')]_+)^2}{2x(\omega')}$$

存在且等于 (4.417) 的左端. 因此, 序列 $\{\bar{\omega}_n\}$ 可在 Γ_ω 中选取. 再一次像定理 4.149 的证明那样, 可以得到 $\bar{t}_n := -2x(\bar{\omega}_n)/h(\bar{\omega}_n)$ 是正数且满足 $\bar{t}_n \downarrow 0$. 因为 Γ_ω 是连通的, 由连续性推证得到, 对 $t_n \in \sigma$ 及充分大的 n, 存在 $\omega_n \in \Gamma_\omega$, 满足 $t_n = -2x(\omega_n)/h(\omega_n)$ 与 $\omega_n \to \omega$. 因此有

$$\begin{aligned} \mathrm{e} - \liminf_{n\uparrow\infty} \frac{x(\omega)+t_n h(\omega)}{t_n^2/2} &\leqslant \liminf_{n\uparrow\infty} \frac{x(\omega_n)+t_n h(\omega_n)}{t_n^2/2} \\ &= \lim_{n\uparrow\infty} \frac{-h(\omega_n)}{2x(\omega_n)} = -\theta(\omega), \end{aligned}$$

从而 $\theta(\omega) \leqslant \tau_{x,h}^{\sigma}(\omega)$. 对任何 $\sigma \in \Sigma$, $\tau_{x,h}(\omega)=\tau_{x,h}^{\sigma}(\omega)$, 因此 $\tau_{x,h}(\omega)=\tau_{x,h}^{*}(\omega)$, 这完成了证明. □

注意到, 条件 (4.417) 只是意味着可以在 Ω 中构造连续的曲线, 它从点 ω 出发, 满足 (4.417) 中左端的上极限可沿这一曲线上取到.

例 4.152　令 Ω 是 $\mathbb{R}^n$ 的紧致子集, $K := C_+(\Omega)$, 令 $x \in C_+(\Omega)$ 满足 $\Delta(x)=\{\omega_0\}$ 是单点集, 其中 ω_0 是 Ω 的内部点. 令 $h \in T_K(x)$ 满足 $h(\omega_0)=0$, 且设 $x(\cdot)$ 在 ω_0 处是二次连续可微的, 其 Hesse 阵 $\nabla^2 x(\omega_0)$ 是正定的, $h(\cdot)$ 在 ω_0 处是连续可微的. 则

$$\theta(\omega_0)=\sup_{\eta} \frac{[\eta^{\mathrm{T}}\nabla h(\omega_0)]^2}{\eta^{\mathrm{T}}\nabla^2 x(\omega_0)\eta} = \nabla h(\omega_0)^{\mathrm{T}}[\nabla^2 x(\omega_0)]^{-1}\nabla h(\omega_0),$$

上述上确界在 $\bar{\eta} := [\nabla^2 x(\omega_0)]^{-1}\nabla h(\omega_0)$ 处取到, 于是得到

$$T_K^{i,2}(x,h)=\{z \in C(\Omega) : z(\omega_0) \geqslant \nabla h(\omega_0)^{\mathrm{T}}[\nabla^2 x(\omega_0)]^{-1}\nabla h(\omega_0)\}.$$

进一步, 条件 (4.417) 在 ω_0 处成立, 相应的集合为

$$\Gamma_{\omega_0} := \{\omega \in \Omega : \omega=\omega_0+t\bar{\eta},\ t \in \mathbb{R}_+\},$$

因此有 $T_K^{i,2}(x,h)=T_K^2(x,h)$.

4.10.2 最优值函数的二阶导数

这一节讨论定义在连续函数空间的最优值函数的二阶导数的计算. 令 Φ 是非空的紧致度量空间. 考虑参数化问题

$$(\mathrm{P}_u) \qquad \min_{x\in\Phi} u(x), \tag{4.418}$$

其中 u 视为在泛函空间 $U := C(\Phi)$ 内变化的参数, 即这一问题的最优值函数 $v(u)$ 可视为元素 $u \in C(\Phi)$ 的函数. 为应用 4.10.1 节的结果, 考虑最大值函数 $\bar{v}(u) := \sup\limits_{x\in\Phi} u(x)$ 也是适合的. 显然, 简单关系 $v(u) = -\bar{v}(-u)$ 在极小与极大函数间成立.

极大值函数 $\bar{v} : C(\Phi) \to \mathbb{R}$ 是正常凸的模为 1 的 Lipschitz 连续函数. 它的上图 $\mathrm{epi}\bar{v} \subset C(\Phi) \times \mathbb{R}$ 可以表示为下述形式

$$\mathrm{epi}\bar{v} = \{(u,\alpha) : \alpha \geqslant u(x),\ \forall x \in \Phi\}.$$

注意到, 由于 $\bar{v}(\cdot)$ 是 Lipschitz 连续的, 它的一阶方向导数是有限的且 Lipschitz 连续的, 一阶与二阶方向导数与相对应的方向上图导数相同. 由命题 2.58 有

$$T_{\mathrm{epi}\bar{v}}(u, \bar{v}(u)) = \mathrm{epi}\bar{v}'(u, \cdot), \tag{4.419}$$

且根据 (3.85) 和 (3.86),

$$\begin{aligned}
T^{i,2}_{\mathrm{epi}\bar{v}}((u,\bar{v}(u)),(\eta,\bar{v}'(u,\eta))) &= \mathrm{epi}\bar{v}''_{+}(u;\eta,\cdot),\\
T^{2}_{\mathrm{epi}\bar{v}}((u,\bar{v}(u)),(\eta,\bar{v}'(u,\eta))) &= \mathrm{epi}\bar{v}''_{-}(u;\eta,\cdot).
\end{aligned}$$

考虑从 $C(\Phi) \times \mathbb{R}$ 到 $C(\Phi)$ 的映射 $G(u,\alpha) := \alpha - u$. 不难得到 $\mathrm{epi}\bar{v} = G^{-1}(C_{+}(\Phi))$. 由于 G 是线性的连续的且映上的, $\mathrm{epi}\bar{v}$ 的二阶切集是 $C_{+}(\Phi)$ 的相应的二阶切集合的逆映像. 更确切地, 有

$$\begin{aligned}
r \geqslant \bar{v}''_{+}(u;\eta,\varrho) &\Leftrightarrow (\varrho, r) \in T^{i,2}_{\mathrm{epi}\bar{v}}[(u,\bar{v}(u)),(\eta,\bar{v}'(u,\eta))]\\
&\Leftrightarrow G(\varrho,r) \in T^{i,2}_{C_{+}(\Phi)}[G(u,\bar{v}(u)),G(\eta,\bar{v}'(u,\eta))]\\
&\Leftrightarrow r-\varrho \in T^{i,2}_{C_{+}(\Phi)}[G(u,\bar{v}(u)),G(\eta,\bar{v}'(u,\eta))].
\end{aligned}$$

于是得

$$\bar{v}''_{+}(u;\eta,\varrho) = \inf\{r : r-\varrho \in T^{i,2}_{C_{+}(\Phi)}[G(u,\bar{v}(u)),G(\eta,\bar{v}'(u,\eta))]\}, \tag{4.420}$$

类似地

$$\bar{v}''_{-}(u;\eta,\varrho) = \inf\{r : r-\varrho \in T^{2}_{C_{+}(\Phi)}[G(u,\bar{v}(u)),G(\eta,\bar{v}'(u,\eta))]\}. \tag{4.421}$$

用定理 4.147 中给出的 $T^{i,2,\sigma}_{C_+(\Phi)}$ 与 $T^{i,2}_{C_+(\Phi)}$ 的表达式, 得到极大值函数 $\bar{v}(\cdot)$ 的二阶方向导数的下述公式. 由 Danskin 定理 4.13,

$$\bar{v}'(u,\eta) = \sup_{x\in\Delta(\bar{u})} \eta(x), \tag{4.422}$$

其中 $\Delta(\bar{u})$ 是在 (4.405) 中定义的函数 $\bar{u}(\cdot) := v(u) - u(\cdot)$ 的接触点的集合. 注意到 $\bar{u}\in C_+(\Phi)$, $\Delta(\bar{u})$ 与 $u(\cdot)$ 在 Φ 上的最大点集合相同. 极大值函数的一阶方向导数的公式 (4.442) 也可以用 (4.419) 通过 $C_+(\Phi)$ 的切锥的相应公式推得.

定理 4.153 设 Φ 是紧致的度量空间, $\bar{v}: C(\Phi)\to\mathbb{R}$ 是相应的最大值函数. 则对任何 u, η, $\varrho\in C(\Phi)$,

$$\bar{v}''_+(u;\eta,\varrho) = \sup_{x\in\Phi}\{\varrho(x)+\tau_{\bar{u},\bar{\eta}}(x)\}, \tag{4.423}$$

$$\bar{v}''_-(u;\eta,\varrho) = \inf_{\sigma\in\Sigma}\sup_{x\in\Phi}\{\varrho(x)+\tau^{\sigma}_{\bar{u},\bar{\eta}}(x)\}, \tag{4.424}$$

其中 $\bar{u}(\cdot) := \bar{v}(u) - u(\cdot)$, $\bar{\eta}(\cdot) := \bar{v}'(u,\eta) - \eta(\cdot)$.

证明 公式 (4.423) 由内二阶切集公式 (4.410) 与 (4.420) 得到, 类似地, (4.424) 由 (4.409) 得到. □

注 4.154 注意, 对所有的 $x\in\Delta(\bar{u})$, 有 $\bar{\eta}(x)\geqslant 0$, 因此 $\bar{\eta}\in T_{C_+(\Phi)}(\bar{u})$. 注意到, 由定理 4.149 中的公式 (4.414), 对所有的 $x\notin\Delta_1(\bar{u},\bar{\eta})$, 有 $\tau_{\bar{u},\bar{\eta}}(x) = -\infty$, 其中

$$\Delta_1(\bar{u},\bar{\eta}) := \{x\in\Delta(\bar{u}) : \bar{\eta}(x) = 0\}. \tag{4.425}$$

因此, 只需要在 (4.423) 的右端或 (4.424) 的右端对 $x\in\Delta_1(\bar{u},\bar{\eta})$ 取上确界.

这一节的余下部分处理极小值函数 $v(u)$ 而不是极大值函数 $\bar{v}(u)$. 所以, 对极小值函数写出定理 4.153 中相对应的公式是合适的, 即对任何 $u,\eta,\varrho\in C(\Phi)$, 有

$$v''_-(u;\eta,\varrho) = \inf_{x\in\Phi}\{\varrho(x)-\tau_{u^*,\eta^*}(x)\}, \tag{4.426}$$

$$v''_+(u;\eta,\varrho) = \sup_{\sigma\in\Sigma}\inf_{x\in\Phi}\{\varrho(x)-\tau^{\sigma}_{u^*,\eta^*}(x)\}, \tag{4.427}$$

其中 $u^*(\cdot) := u(\cdot) - v(u)$, $\eta^*(\cdot) := \eta(\cdot) - v'(u,\eta)$.

由 Danskin 定理, 方向导数 $v'(u,\eta)$ 可以表示为形式

$$v'(u,\eta) = \inf_{x\in\mathcal{S}(u)} \eta(x), \tag{4.428}$$

其中

$$\mathcal{S}(u) := \operatorname*{argmin}_{x\in\Phi}\ u(x).$$

由定理 4.149 中的公式 (4.414), 有

$$\tau_{u^*,\eta^*}(x) = \begin{cases} 0, & \text{若 } x \in \operatorname{int}(\mathcal{S}(u)) \text{ 且 } \eta(x) = v'(u,\eta), \\ \vartheta(x), & \text{若 } x \in \operatorname{bdr}(\mathcal{S}(u)) \text{ 且 } \eta(x) = v'(u,\eta), \\ -\infty, & \text{否则}, \end{cases} \tag{4.429}$$

其中

$$\vartheta(x) := \limsup_{\substack{x' \to x \\ u(x') > v(u)}} \frac{([v'(u,\eta) - \eta(x')]_+)^2}{2[u(x') - v(u)]}. \tag{4.430}$$

注意到, 集合 $\mathcal{S}(u)$ 是闭的, 因此 $\mathcal{S}(u) = \operatorname{int}(\mathcal{S}(u)) \cup \operatorname{bdr}(\mathcal{S}(u))$. 又注意到 $\vartheta(x)$ 总是非负的. 因此, 只需在 (4.426) 的右端在下述集合上取下确界

$$\mathcal{S}_1(u,\eta) := \{x \in \operatorname{bdr}(\mathcal{S}(u)) : \eta(x) = v'(u,\eta)\}, \tag{4.431}$$

如果此集合是非空的. 结合给出确保相应的内二阶切集与外二阶切集相等的条件的命题 4.151, 可推出下述结果.

定理 4.155 *设 Φ 是紧致的度量空间, $v : C(\Phi) \to \mathbb{R}$ 是极小值函数. 则对任何 $u, \eta, \varrho \in C(\Phi)$, 有*

$$v''_-(u;\eta,\varrho) = \begin{cases} \inf\limits_{x \in \mathcal{S}_1(u,\eta)} \{\varrho(x) - \vartheta(x)\}, & \text{若} \mathcal{S}_1(x,u) \neq \varnothing, \\ 0, & \text{否则}, \end{cases} \tag{4.432}$$

其中 $\vartheta(x)$ 与 $\mathcal{S}_1(u,\eta)$ 分别由 (4.430) 与 (4.431) 定义. 进一步, $v''_-(u;\eta,\cdot) = v''_+(u;\eta,\cdot)$, 即在下述条件成立的前提下, $v(\cdot)$ 在 u 处沿方向 η 二阶方向可微成立: 对任何 $x_0 \in \mathcal{S}_1(u,\eta)$, 存在连续路径 $\bar{x} : [0,1] \to \Phi$ 满足 $\bar{x}(0) = x_0$, 对所有的 $t \in (0,1]$, $u(\bar{x}(t)) > v(u)$, 且 (4.430) 式右端的 limsup 沿此路径取到极限, 即

$$\limsup_{\substack{x \to x_0 \\ u(x) > v(u)}} \frac{([v'(u,\eta) - \eta(x)]_+)^2}{2[u(x) - v(u)]} = \liminf_{t \downarrow 0} \frac{([v'(u,\eta) - \eta(\bar{x}(t))]_+)^2}{2[u(\bar{x}(t)) - v(u)]}. \tag{4.433}$$

遗憾的是, 不总是很容易计算 $\vartheta(x)$ 在给定点 $x_0 \in \mathcal{S}_1(u,\eta)$ 的值. 例如, 设 Φ 是赋范向量空间的紧子集, 一阶增长条件在 x_0 处成立, 即存在某一 $c > 0$, 对 x_0 的邻域中的所有 $x \in \Phi$, 有 $u(x) \geqslant u(x_0) + c\|x - x_0\|$. 这当然意味着 x_0 是 $\mathcal{S}(u)$ 的孤立点. 进一步, 设 $\eta(\cdot)$ 在 x_0 处是可微的. 则 (4.433) 左端比值中的分子是 $O(\|x - x_0\|^2)$ 阶的, 而分母至少以 $2c\|x - x_0\|$ 的速率增长. 结果, 此种情况有 $\vartheta(x_0) = 0$. 也可以在二阶增长条件成立的点处计算 $\vartheta(x)$. 然而, 这就比较具有技巧性了. 下一节以一种直接的方式对最优值函数实现这样的计算, 还将研究函数 $v(u)$ 的二阶正则性.

4.10.3 泛函空间的二阶展开

4.10.2 节已经讨论了定义在空间 $C(\Phi)$ 上的最优值函数的二阶导数. 这一节将以一种直接的方式研究最优值函数的二阶展开及最优解的相应的一阶展开. 设 Φ 是有限维空间的非空的紧致子集, 如 $\Phi \subset \mathbb{R}^n$. 考虑由 (4.418) 定义的相应的极小化问题 (P_u)、最优值函数 $v(u) := \inf_{x\in\Phi} u(x)$ 与最优解 $\bar{x}(u)$, 它们可视为在合适的泛函空间 U 中的元素 u 的函数.

4.10.2 节对泛函空间 $U := C(\Phi)$ 的情形给出分析. 然而, 由于需要所涉及函数的一定的 Lipschitz 性质, 空间 $C(\Phi)$ 就显得太大. 因此, 考虑 Lipschitz 连续函数 $\psi : \Phi \to \mathbb{R}$ 构成的空间 $W^{1,\infty}(\Phi)$, 赋予范数

$$\|\psi\| := \sup_{x\in\Phi} |\psi(x)| + \sup\left\{\frac{|\psi(x') - \psi(x)|}{\|x' - x\|} : x, x' \in \Phi, x' \neq x\right\}. \tag{4.434}$$

式 (4.434) 右端第二项代表函数 ψ 的 Lipschitz 常数, 若集合 Φ 是具有非空内部的凸集合, 当 $\nabla\psi(x)$ 存在时, 它可以表示为 $\|\nabla\psi(x)\|$ 对所有的 $x \in \Phi$ 取上确界.

现在考虑参数空间 $U := W^{1,\infty}(\Phi)$ 中的下述路径:

$$u_t(\cdot) := f(\cdot) + t\eta(\cdot) + \varepsilon_t(\cdot), \quad t \geqslant 0, \tag{4.435}$$

其中 $f, \eta, \varepsilon_t \in W^{1,\infty}(\Phi)$ 且 $\|\varepsilon_t\| = o(t)$. 用 $\nu(t)$ 记 (P_{u_t}) 的最优值函数, 即 $\nu(t) := v(u_t)$, 或等价地,

$$\nu(t) := \inf_{x\in\Phi} \{f(x) + t\eta(x) + \varepsilon_t(x)\}. \tag{4.436}$$

令 x_0 是 $f(\cdot)$ 在 Φ 上的极小点, 即 x_0 是非扰动问题 (P) 的最优解. 由 Danskin 定理 4.13, 有 $\nu(t)$ 在 $t = 0$ 处是可微的且 $d\nu(0)/dt = \eta(x_0)$, 若假设最优解 x_0 是唯一的.

设 f 定义在集合 Φ 的邻域上, f 在点 x_0 处是 Hadamard 方向可微的. 若函数 $f(\cdot)$ 在 $x_0 \in \mathbb{R}^n$ 的邻域内为 Lipschitz 连续的, 则在 x_0 处的 Hadamard 意义的方向可微性可由它在 x_0 处的方向可微性得到. 由假设 $f \in W^{1,\infty}(\Phi)$, f 在 Φ 上是 Lipschitz 连续的. 然而, 如果将 x_0 视为 R^n 中的点, 这并不意味着 f 在 x_0 的邻域内是 Lipschitz 连续的.

设 (4.435) 中考虑的函数 $\eta(\cdot)$ 与 $\varepsilon_t(\cdot)$ 也定义在集合 Φ 的邻域上, $\eta(\cdot)$ 在 x_0 处是 Hadamard 方向可微的, 且 f 在 x_0 处是二阶方向可微的. 由二阶上图正则函数的定义 3.94, 即函数在 x_0 处沿方向 h 是二阶上图正则的, 当满足 $t \downarrow 0$ 时, $tw(t) \to 0$ 的路径 $w(\cdot) : \mathbb{R}_+ \to \mathbb{R}^n$ 有

$$f\left(x_0 + th + \frac{1}{2}t^2 w(t)\right) \geqslant f(x_0) + tf'(x_0, h) + \frac{1}{2}t^2 f''(x_0; h, w(t)) + o(t^2). \tag{4.437}$$

由命题 3.95 可得, 若 f 在 $x_0 \in \mathbb{R}^n$ 的邻域内是 Lipschitz 连续的, 则 f 在 x_0 处沿方向 h 是二阶上图正则的充分必要条件是 f 在 x_0 处沿方向 h 是外二阶正则的. 一阶必要条件可以表示为下述形式 (如命题 3.99)

$$f'(x_0, h) \geqslant 0, \quad \forall\, h \in T_\Phi(x_0), \tag{4.438}$$

相应的临界锥可以表示为

$$C(x_0) := \{h \in T_\Phi(x_0) : f'(x_0, h) = 0\}. \tag{4.439}$$

由命题 3.105 可得, 若 f 在 x_0 处沿每一方向 $h \in C(x_0)$ 是二阶上图正则的, 则二阶增长条件在 x_0 处成立的充分必要条件是

$$\inf_{w \in \mathbb{R}^n} f''(x_0; h, w) > 0, \quad \forall\, h \in C(x_0) \backslash \{0\}. \tag{4.440}$$

现在推导最优值函数与最优解的下述展开. 若集合 Φ 在 x_0 处沿方向 h 是二阶正则的, 则可得到在 x_0 处沿方向 h 的外二阶切集与内二阶切集是重合的.

定理 4.156 令 x_0 是 (P) 的唯一最优解, $\bar{x}(t)$ 是问题 (P_{u_t}) 的 $o(t^2)$ 最优解. 设

(i) 函数 $f, \eta : \mathbb{R}^n \to \mathbb{R}$ 在 x_0 处是 Hadamard 方向可微的.

(ii) f 在 x_0 处是二阶方向可微的.

(iii) f 在 x_0 处沿任何方向 $h \in C(x_0)$ 均是二阶上图正则的.

(iv) 集合 Φ 在 x_0 处沿任何方向 $h \in C(x_0)$ 均是二阶正则的.

(v) 非扰动问题 (P) 在点 x_0 处的二阶增长条件成立.

则

(a) 对 $t \geqslant 0$,

$$\nu(t) = \nu(0) + t\eta(x_0) + \varepsilon_t(x_0) + \frac{1}{2}t^2 \zeta_{f,\eta}(x_0) + o(t^2), \tag{4.441}$$

其中

$$\zeta_{f,\eta}(x) := \inf_{h \in C(x)} \{2\eta'(x, h) + \inf_{w \in T^2_\Phi(x,h)} f''(x\,; h, w)\}. \tag{4.442}$$

(b) $\bar{x}(t)$ 在 x_0 处是 Lipschitz 稳定的, 当 $t \downarrow 0$ 时, $t^{-1}(\bar{x}(t) - x_0)$ 的任何极限点是下述问题的最优解

$$\min_{h \in C(x_0)} \{\eta'(x_0, h) + \frac{1}{2} \inf_{w \in T^2_\Phi(x_0,h)} f''(x_0; h, w)\}. \tag{4.443}$$

(c) 若还有, 问题 (4.443) 具有唯一的最优解 $\bar{h}$, 则

$$\bar{x}(t) = x_0 + t\bar{h} + o(t), \quad t \geqslant 0.$$

证明　令 $\bar{x}(t)$ 是 (P_{u_t}) 的 $o(t^2)$ 最优解. 则当 $t\downarrow 0$ 时, $\bar{x}(t)\to x_0$ (见命题 4.4). 进一步, 根据命题 4.32, 由二阶增长条件可得

$$\|\bar{x}(t)-x_0\| = O(\kappa(t))+o(t),$$

其中 $\kappa(t)$ 是函数 $t\eta(\cdot)+\varepsilon_t(\cdot)$ 在 Φ 上的 Lipschitz 常数. 由于 $\eta\in W^{1,\infty}(\Phi)$, 它在 Φ 上是 Lipschitz 连续的, $\varepsilon_t(\cdot)$ 的 Lipschitz 常数是 $o(t)$ 阶的, 有 $\kappa(t)=O(t)$, 因此 $\|\bar{x}(t)-x_0\|=O(t)$.

现在考虑形式为 $x(t):=x_0+th+\dfrac{1}{2}t^2w+o(t^2)$ 的可行路径, 其中 $h\in C(x_0)$ 与 $w\in T^2_\Phi(x_0,h)$. 注意到, 由于 Φ 的二阶正则性成立, 有 $T^{i,2}_\Phi(x_0,h)=T^2_\Phi(x_0,h)$, 因此这样的路径是可选取为可行的, 即由内二阶切集 $T^{i,2}_\Phi(x_0,h)$ 的定义, 则 $x(t)\in\Phi$ 可取得. 由于 $f'(x_0,h)=0$, $\varepsilon_t(\cdot)$ 的 Lipschitz 常数是 $o(t)$ 阶的, 则有

$$\|\varepsilon_t(x(t))-\varepsilon_t(x_0)\| = o(t)\|x(t)-x_0\| = o(t^2), \tag{4.444}$$

得到

$$\begin{aligned}\nu(t)&\leqslant f(x(t))+t\eta(x(t))+\varepsilon_t(x(t))\\&= f(x_0)+\frac{1}{2}t^2f''(x_0;h,w)+t\eta(x_0)+t^2\eta'(x_0,h)+\varepsilon_t(x_0)+o(t^2).\end{aligned}$$

对任何 $h\in C(x_0)$ 与 $w\in T^2_\Phi(x_0,h)$, 有

$$\limsup_{t\downarrow 0}\frac{\nu(t)-\nu(0)-t\eta(x_0)-\varepsilon_t(x_0)}{\dfrac{1}{2}t^2}\leqslant 2\eta'(x_0,h)+f''(x_0;h,w),$$

因此 $\nu(t)$ 小于或等于 (4.441) 右端.

为了证明另外的不等式, 考虑序列 $t_n\downarrow 0$ 与 $x_n:=\bar{x}(t_n)$. 因为 $\|x_n-x_0\|=O(t_n)$, 如有必要可取一子序列, 不妨设 $t_n^{-1}(x_n-x_0)$ 收敛到向量 h. 有 $h\in T_\Phi(x_0)$, 因为

$$\nu(t)=f(\bar{x}(t))+t\eta(\bar{x}(t))+\varepsilon_t(\bar{x}(t))+o(t^2)=\nu(0)+tf'(x_0,h)+t\eta(x_0)+o(t)$$

与 $d\nu(0)/dt=\eta(x_0)$, 得到 $h\in C(x_0)$. 可以表示

$$x_n=x_0+t_nh+\frac{1}{2}t_n^2w_n,$$

其中 w_n 满足 $t_nw_n\to 0$. 由 Φ 的二阶正则性, 存在 $w_n'\in T^2_\Phi(x_0,h)$ 满足 $\|w_n-w_n'\|\to 0$. 由 (4.437) 得到

$$f\left(x_0+t_nh+\frac{1}{2}t_n^2w_n'\right)\geqslant f(x_0)+t_nf'(x_0,h)+\frac{1}{2}t_n^2f''(x_0;h,w_n')+o(t_n^2).$$

所以, 置 $\varepsilon_n := \varepsilon_{t_n}(x_0)$, 由 (4.444) 得 $\|\varepsilon_{t_n}(x_0) - \varepsilon_n\| = o(t_n^2)$, 所以

$$
\begin{aligned}
\nu(t_n) &= f(x_n) + t_n\eta(x_n) + \varepsilon_{t_n}(x_n) + o(t_n^2) \\
&\geqslant f\left(x_0 + t_n h + \frac{1}{2}t_n^2 w_n'\right) + t_n\eta\left(x_0 + t_n h + \frac{1}{2}t_n^2 w_n'\right) + \varepsilon_n + o(t_n^2) \\
&\geqslant f(x_0) + \frac{1}{2}t_n^2 f''(x_0; h, w_n') + t_n\eta(x_0) + t_n^2\eta'(x_0, h) + \varepsilon_n + o(t_n^2) \\
&\geqslant \nu(0) + t_n\eta(x_0) + \frac{1}{2}t_n^2[2\eta'(x_0, h) + \inf_{w\in T_\Phi^2(x_0,h)} f''(x_0; h, w)] + \varepsilon_n + o(t_n^2).
\end{aligned}
$$

这证得另外一不等式, 因此 (4.441) 成立, 还得到 h 是问题 (4.443) 的最优解.

最后的结论 (c) 由结论 (b) 可得到, 这就完成定理的证明. □

注意, 如果 $\varepsilon_t(\cdot)$ 的 Lipschitz 常数是 $o(t)$ 阶的, 定理 4.156 的结论 (b) 与 (c) 不依赖于余项 ε_t.

我们可以将项 $\varepsilon_t(\cdot)$ 写成 $\varepsilon_t(\cdot) = \frac{1}{2}t^2\varrho_t(\cdot)$, 其中 $\varrho_t(\cdot) = 2t^{-2}\varepsilon_t(\cdot)$, 则由公式 (4.441), 在定理 4.156 的假设之下, 最小值函数 $v(u)$ 在 f 处沿方向 η 是二阶方向可微的与二阶正则的, 且

$$
v''(f; \eta, \varrho) = \varrho(x_0) + \zeta_{f,\eta}(x_0). \tag{4.445}
$$

进一步, 若集合 $\mathcal{S}(f) := \operatorname*{argmin}_{x\in\Phi} f(x)$ 是有限的, 且定理 4.156 的假设在 $\mathcal{S}(f)$ 的每一点均成立, 与定理 4.155 的公式 (4.432) 相比较, 可得到下述结果.

定理 4.157 设集合 $\mathcal{S}(f)$ 是有限的, 定理 4.156 的假设在每一点 $x \in \mathcal{S}(f)$ 处成立, 则极小值函数在 f 处沿方向 η 是二阶方向可微的与二阶正则的, 且

$$
v''(f; \eta, \varrho) = \min_{x\in\mathcal{S}_1(f,\eta)} \{\varrho(x) + \zeta_{f,\eta}(x)\}, \tag{4.446}
$$

其中 $\mathcal{S}_1(f, \eta) := \operatorname*{argmin}_{x\in\mathcal{S}(f)} \eta(x)$, $\zeta_{f,\eta}(x)$ 由 (4.442) 定义.

现在设 f 是二阶连续可微的, 则 f 是二阶上图正则的, 且

$$
f''(x_0; h, w) = Df(x)w + D^2 f(x)(h, h), \tag{4.447}
$$

因此函数 $\zeta_{f,\eta}(\cdot)$ 取为下述形式

$$
\zeta_{f,\eta}(x) = \inf_{h\in C(x)} \{2\eta'(x, h) + D^2 f(x)(h, h) + \inf_{w\in T_\Phi^2(x,h)} Df(x)w\}. \tag{4.448}
$$

在集合 Φ 是多面集的情形, 它是二阶正则的, 有

$$
T_\Phi^2(x, h) = T_{T_\Phi(x)}(h) = \operatorname{cl}\{T_\Phi(x) + [\![h]\!]\}.
$$

因为对任何 $h \in C(x)$ 与 $w \in T_\Phi(x)$, $Df(x)h = 0$, $Df(x)w \geqslant 0$, 则 (4.448) 右端的第三项在此种情形下消失了.

现在设集合 Φ 由 $\Phi := G^{-1}(K)$ 给出, 其中 K 是 Banach 空间 Y 的闭凸子集且 $G : \mathbb{R}^n \to Y$ 是二阶连续可微的映射, 在点 $x \in \mathcal{S}(f)$ 处的 Robinson 约束规范成立, 则

$$T_\Phi(x) = DG^{-1}[T_K(G(x))], \tag{4.449}$$

且 (见命题 3.33 与 (3.59))

$$T_\Phi^2(x,h) = DG(x)^{-1}[T_K^2(G(x), DG(x)h) - D^2G(x)(h,h)]. \tag{4.450}$$

由 (4.450), (4.448) 右端第三项由下述问题的最优值给出

$$\min_w \ Df(x)w \quad \text{s. t.} \quad DG(x)w + D^2G(x)(h,h) \in \mathcal{T}_K^2(h), \tag{4.451}$$

其中

$$\mathcal{T}_K^2(h) := T_K^2(G(x), DG(x)h).$$

在 Robinson 约束规范之下, 问题 (4.451) 的最优值与其对偶问题的最优值重合 (见 3.2.2 节中对偶问题 (3.104) 的推导):

$$\max_{\lambda \in \Lambda(x)} \{\langle \lambda, D^2G(x)(h,h)\rangle - \sigma(\lambda, \mathcal{T}_K^2(h))\}, \tag{4.452}$$

其中 $\Lambda(x)$ 是问题 (P) 在 x 处的 Lagrange 乘子集. 结果, 在此种情况下有

$$\zeta_{f,\eta}(x) = \inf_{h \in C(x)} \sup_{\lambda \in \Lambda(x)} \{2\eta'(x,h) + D_{xx}^2 L(x,\lambda)(h,h) - \sigma(\lambda, \mathcal{T}_K^2(h))\}. \tag{4.453}$$

若集合 K 是多面集, 则上述公式中的 sigma 项消失.

极小值函数 $v(u)$ 的二阶方向导数的公式 (4.446) 可以用来推导泛函空间的非负值函数锥的二阶切集的表达式, 从而可以用来检验其二阶正则性, 即考虑锥 $K := W_+^{1,\infty}(\Phi)$, 其中

$$W_+^{1,\infty}(\Phi) := \{u \in W^{1,\infty}(\Phi) : u(x) \geqslant 0, \forall x \in \Phi\}. \tag{4.454}$$

这一锥也可以表示为

$$W_+^{1,\infty}(\Phi) := \{u \in W^{1,\infty}(\Phi) : v(u) \geqslant 0\}. \tag{4.455}$$

注意 Slater 条件成立, 即存在 $\bar{u} \in W^{1,\infty}(\Phi)$ 满足 $v(\bar{u}) > 0$(如取 $\bar{u}(x) = 1$ 对所有 $x \in \Phi$ 成立), 且最优值函数 $v(u)$ 是 Lipschitz 连续的, 在 $W^{1,\infty}(\Phi)$ 上是凹函数.

考虑函数 $f\in W^{1,\infty}_+(\Phi)$, 记

$$\Delta(f):=\{x\in\Phi:f(x)=0\} \tag{4.456}$$

是起作用点的集合. 若 $v(f)>0$, 因而 $\Delta(f)$ 是空集, 则 f 是 $W^{1,\infty}_+(\Phi)$ 的内点, 此时 $T_K(f)=W^{1,\infty}(\Phi)$. 所以, 设集合 $\Delta(f)$ 是非空的. 此种情形下, $\Delta(f)$ 与 $f(x)$ 在 Φ 上的极小值点集是重合的, 即 $\Delta(f)=\mathcal{S}(f)$, 且

$$T_K(f)=\{u\in W^{1,\infty}(\Phi):u(x)\geqslant 0,\ \forall x\in\Delta(f)\}. \tag{4.457}$$

对 $\eta\in T_K(f)$, 置

$$\Delta_1(f,\eta):=\{x\in\Delta(f):\eta(x)=0\}. \tag{4.458}$$

若 $\Delta_1(f,\eta)$ 是空集, 对任何 $t>0$, 则 $f+t\eta$ 是 $W^{1,\infty}_+(\Phi)$ 是内点且 $T^2_K(f,\eta)$ 与整个空间 $W^{1,\infty}(\Phi)$ 重合. 若 $\Delta_1(f,\eta)$ 是非空的, 则 $\Delta_1(f,\eta)=\mathcal{S}_1(f,\eta)$. 由命题 3.30 与 3.92, 下述结果由定理 4.157 可得到.

命题 4.158 令 $f\in K:=W^{1,\infty}_+(\Phi)$, $\eta\in T_K(f)$, 设集合 $\Delta(f)$ 与 $\Delta_1(f)$ 是非空的有限的, 且对每一 $x_0\in\Delta(f)$, 定理 4.156 的假设 (i)~(v) 成立. 则集合 $W^{1,\infty}_+(\Phi)$ 在 f 处沿方向 η 是二阶正则的, 且

$$T^2_K(f,\eta)=\{u\in W^{1,\infty}(\Phi):u(x)+\zeta_{f,\eta}(x)\geqslant 0,\ \forall x\in\Delta_1(f,\eta)\}, \tag{4.459}$$

其中 $\zeta_{f,\eta}(\cdot)$ 由 (4.442) 定义.

公式 (4.459) 可同定理 4.148 中的公式 (4.410) 相比较.

若所考虑的泛函空间是 $U:=C(\Phi)$, 由相似的推证可以证明, 在定理 4.156 的假设之下, 公式 (4.446) 仍然是成立的. 然而, 此种情况下, $v''(f;\eta,\varrho)$ 关于 $\varrho\in C(\Phi)$ 不是一致的, 如果没有额外的假设, $v(\cdot)$ 与 $C_+(\Phi)$ 的二阶正则性的类似结果不能得到, 这是因为余项 ε_t 以 $C(\Phi)$ 的 sup 范数为 $o(t)$ 阶的这一假设不能保证 $\varepsilon_t(\cdot)$ 的 Lipschitz 常数是 $o(t)$ 阶的. 为克服这一困难, 考虑嵌入映射 $l:W^{1,\infty}(\Phi)\to C(\Phi)$ (即 l 映 $u\in W^{1,\infty}(\Phi)$ 到 $u\in C(\Phi)$), 则下述结论成立.

命题 4.159 令 $f\in K:=C_+(\Phi)$, $\eta\in T_K(f)$, 设集合 $\Delta(f)$ 与 $\Delta_1(f,\eta)$ 是非空有限的, 且对每一 $x_0\in\Delta(f)$, 定理 4.156 中的假设 (i)~(v) 成立. 则集合 $C_+(\Phi)$ 在 f 处沿方向 η 关于嵌入映射 l 是二阶正则的, 且

$$T^2_K(f,\eta)=\{u\in C(\Phi):u(x)+\zeta_{f,\eta}(x)\geqslant 0,\forall x\in\Delta_1(f,\eta)\}, \tag{4.460}$$

其中 $\zeta_{f,\eta}(\cdot)$ 由 (4.442) 定义.

第 5 章　额外的素材及应用

本章研究参数变分不等式 (广义方程), 讨论非线性规划问题、半定规划问题与半无限问题理论的应用. 一般性理论具体应用的不同表现已经在前面的章节里讨论过. 然而, 我们建议对这些应用的某一专题感兴趣的读者先阅读本章的相应小节, 以便对这些专题的最优性条件与扰动理论的威力有一个整体的了解. 另外, 本章还给出了前面章节中没有提及的某些结果.

5.1　变分不等式

这一节讨论变分不等式与所谓的广义方程的扰动理论. 变分不等式与广义方程的灵敏度分析不像最优化问题的扰动分析那样完善. 困难之一在于对偶理论的强有力的工具不能应用到广义方程中. 在最优化问题的情形, 由紧致性推证可以相对容易地建立最优解的存在性, 在广义方程的情形, 验证解的存在性经常不是一个平凡的问题. 这一节, 我们不试图给出变分不等式 (广义方程) 的完整理论, 而是有选择性地给出几个结果.

5.1.1　标准变分不等式

考虑所谓的 (标准的) 变分不等式: 求 $x \in K$ 满足

$$\text{(VI)} \qquad \langle F(x), y-x\rangle \geqslant 0, \quad \forall y \in K, \tag{5.1}$$

其中 K 是 Banach 空间X 的非空闭凸子集, $F : X \to X^*$ 是连续可微的映射. 若 $F(x)$ 可以表示为实值函数 $f : X \to R$ 的导数, 即 $F(x) = Df(x)$, 则变分不等式 (5.1) 等价于 $f(x)$ 在 $x \in K$ 上的极小化问题的 (一阶) 最优性条件. 注意到, F 是二次连续可微函数的导数当且仅当双线性映射 $(h_1, h_2) \to \langle DF(x)h_1, h_2\rangle$ 对所有的 x 均是对称的 (见 2.2.2 节). 若 $F(x) := Ax - b, A \in \mathcal{L}(X, X^*), b \in X^*$, 称 (5.1) 是线性变分不等式. 此种情况, 若空间 X 是自反的, 则 F 可表示为函数的导数的充分必要条件是线性算子 A 是自共轭的, 即 $A = A^*$. 相应的函数 f 可以由 $f(x) := \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$ 给出.

若 $x \in X$ 满足 (5.1), 称 x 是 (5.1) 的解. 用 $\mathcal{S}(\text{VI})$ 记 (5.1) 的所有解的集合. 注意到 $x \in \mathcal{S}(\text{VI})$ 当且仅当 $-F(x) \in N_K(x)$, 或等价地, $0 \in F(x) + N_K(x)$. 不难看到, 集合 $\mathcal{S}(\text{VI})$ 是闭的. 解 $x \in \mathcal{S}(\text{VI})$ 称为是局部唯一的, 若在 x 的某一邻域, 它是唯一

的解.

在解 $x \in \mathcal{S}(\mathrm{VI})$ 处的上述变分不等式的临界锥定义为

$$C(x) := T_K(x) \cap F(x)^{\perp}, \tag{5.2}$$

其中 $F(x)^{\perp} := \{h \in X : \langle F(x), h\rangle = 0\}$. 若 F 可表示为函数 f 的导数, 则临界锥的定义与第 3 章给出的定义重合. 与解 $x \in \mathcal{S}(\mathrm{VI})$(连同 $DF(x)$) 相关的二次型定义为

$$Q_x(h) := \langle DF(x)h, h\rangle.$$

对于多维问题而言, 上述二次型经常是 Legendre 形式. 此种情况, 理论比较完善. 若空间 X 是有限维的, 则任何二次函数均是 Legendre 形式.

算子 F 在 K 上称为是单调的, 若对所有的 $x, y \in K$,

$$\langle F(y) - F(x), y - x\rangle \geqslant 0.$$

若 $K = X$, 上述不等式成立, 称 F 是单调的算子. 例如, 若 F 可表示为凸函数的导数, 则是单调的. 若上述不等式当 $x \neq y$ 时是严格的, 称 F 在 K 上是严格单调的. 由中值定理得

$$F(y) - F(x) = \int_0^1 DF(x + th)h\mathrm{d}t,$$

其中 $h := y - x$, 则 F 是单调的当且仅当对每一 $x \in X$, 相应的二次型 $Q_x(\cdot)$ 是非负的, 即对所有的 $h \in X, Q_x(h) \geqslant 0$. 进一步, 若对所有的 $x \in K$ 与非零的 $h \in \mathrm{Sp}(K - K)$, 则 F 在 K 上是严格单调的.

命题 5.1 *设算子 F 在 K 上是单调的, 则下述结论成立:*

(i) 点 $x \in K$ 是变分不等式 (5.1) 的解的充分必要条件是

$$\langle F(y), y - x\rangle \geqslant 0, \quad \forall y \in K. \tag{5.3}$$

(ii) 问题 (5.1) 的解集 $\mathcal{S}(\mathrm{VI})$ 是凸的, 且若 $x \in \mathcal{S}(\mathrm{VI})$, 则 $\mathcal{S}(\mathrm{VI}) \subset x + C(x)$.

(iii) 进一步, 若 F 在 K 上是严格单调的, 则 (5.1) 至多有一解.

证明 (i) 设 $x \in \mathcal{S}(\mathrm{VI})$. 由于 F 是单调的, 有

$$\langle F(y), y - x\rangle \geqslant \langle F(x), y - x\rangle \geqslant 0, \quad \forall y \in K.$$

相反地, 令 $x \in K$ 满足 (5.3). 置 $y_t := x + t(y - x)$. 则

$$0 \leqslant \lim_{t \downarrow 0} \frac{\langle F(y_t), y_t - x\rangle}{t} = \langle F(x), y - x\rangle.$$

所以, $x \in \mathcal{S}(\mathrm{VI})$.

(ii) 令 x_1 与 x_2 是两个解. 则

$$\langle F(x_1), x_2 - x_1\rangle \geqslant 0 \quad \text{且} \quad \langle F(x_2), x_1 - x_2\rangle \geqslant 0. \tag{5.4}$$

因为 F 在 K 上是单调的, 上述两个不等式的和是非正的. 可以得到

$$\langle F(x_1), x_2 - x_1\rangle = \langle F(x_2), x_1 - x_2\rangle = 0. \tag{5.5}$$

因为 $x_2 - x_1 \in T_K(x)$, 可推出 $\mathcal{S}(\mathrm{VI}) \subset x_1 + C(x_1)$

现在证 $\mathcal{S}(\mathrm{VI})$ 是凸的. 再次设 x_1 与 x_2 是两个解. 令 $t \in (0,1), x_t := tx_1 + (1-t)x_2$, 则对任何 $y \in K$,

$$\langle F(y), y - x_t\rangle = (1-t)\langle F(y), y - x_2\rangle + t\langle F(y), y - x_1\rangle \geqslant 0.$$

结合 (i), 证得 $x_t \in \mathcal{S}(\mathrm{VI})$.

(iii) 若存在两个不同的解 x_1 与 x_2, 则可得 $\langle F(x_2) - F(x_1), x_2 - x_1\rangle > 0$ 与 (5.5) 矛盾. □

现在考虑扰动的变分不等式: 求 $x \in K$ 满足

$$(\mathrm{VI}_u) \qquad \langle F(x,u), y - x\rangle \geqslant 0, \quad \forall y \in K, \tag{5.6}$$

它以 $u \in U$ 为参数, 其中 U 是 Banach 空间, $F : X \times U \to X^*$ 是连续可微的映射. 对给定的参数向量值 u_0, 设 $F(\cdot, u_0)$ 与非扰动算子 $F(\cdot)$ 重合, x_0 是非扰动变分不等式 (5.1) 的解. 用 $\mathcal{S}(u)$ 记 (5.6) 的解集, 尤其 $\mathcal{S}(u_0) = \mathcal{S}(\mathrm{VI})$, 用

$$Q_0(h) := \langle D_x F(x_0, u_0)h, h\rangle$$

记与 $D_x F(x_0, u_0)$ 相联系的二次型.

考虑下述条件: 存在常数 $\beta > 0$ 与 $\eta > 0$, 满足

$$Q_0(h) \geqslant \beta\|h\|^2, \quad \forall h \in \widehat{C}_\eta(x_0), \tag{5.7}$$

其中 $\widehat{C}_\eta(x_0)$ 是近似临界锥, 定义为

$$\widehat{C}_\eta(x_0) := \{h \in T_K(x_0) : \langle F(x_0), h\rangle \leqslant \eta\|h\|\}. \tag{5.8}$$

注意到, 因为 $x_0 \in \mathcal{S}(\mathrm{VI})$, 对任意 $h \in T_K(x_0)$ 有 $\langle F(x_0), h\rangle \geqslant 0$. 还注意到, 若空间 X 是自反的且 Q_0 是 Legendre 形式, 则由引理 3.75 可得, 上述条件 (5.7) 等价于下述条件

$$Q_0(h) > 0, \quad \forall h \in C(x_0)\backslash\{0\}. \tag{5.9}$$

命题 5.2 (i) 若条件 (5.7) 满足, 则对 x_0 的邻域中的所有 $\bar{x}(u) \in \mathcal{S}(u)$, 有

$$\|\bar{x}(u) - x_0\| = O(\|u - u_0\|). \tag{5.10}$$

(ii) 若存在常数 $\beta > 0$ 与 $\eta > 0$ 满足

$$Q_0(h) \geqslant \beta\|h\|^2, \quad \forall h \in \mathrm{Sp}(T_K(x_0)), \quad |\langle F(x_0, u_0), h\rangle| \leqslant \eta\|h\|. \tag{5.11}$$

则存在 $\kappa > 0$, 满足对 u_0 的邻域中所有的 u_1 与 u_2 与 x_0 的邻域中 $\bar{x}(u_1) \in \mathcal{S}(u_1)$ 与 $\bar{x}(u_2) \in \mathcal{S}(u_2)$, 下述不等式成立

$$\|\bar{x}(u_1) - \bar{x}(u_2)\| \leqslant \kappa\|u_1 - u_2\|. \tag{5.12}$$

证明 (i) 设 $u_n \to u_0, x_n \in \mathcal{S}(u_n)$ 满足 $x_n \to x_0$. 置 $\alpha_n := \|x_n - x_0\|, h_n := \alpha_n^{-1}(x_n - x_0)$, 有 $x_n = x_0 + \alpha_n h_n, \|h_n\| = 1$. 要证 $\alpha_n = O(\|u_n - u_0\|)$. 假设不真, 即 $\alpha_n^{-1}(u_n - u_0) \to 0$. 因为 $h_n \in T_K(x_0)$ 且

$$\langle F(x_0, u_0), h_n\rangle \geqslant 0 \quad 且 \quad \langle F(x_n, u_n), -h_n\rangle \geqslant 0,$$

得到对所有充分大的 n, 有 $h_n \in \hat{C}_\eta(x_0)$. 将上述两个不等式相加, 可以得到

$$\langle F(x_n, u_n) - F(x_0, u_0), h_n\rangle \leqslant 0,$$

因此, 由 F 的一阶展式得

$$\alpha_n\langle D_x F(x_0, u_0)h_n, h_n\rangle \leqslant o(\alpha_n).$$

于是, 对充分大的 n, 有 $Q_0(h_n) < \beta$. 这与条件 (5.7) 矛盾, 因此完成了 (i) 的证明.

(ii) 假设 (5.12) 不成立, 则存在收敛到 u_0 的序列 u_n 与 u_n' 以及与它们相联系的 $x_n \in \mathcal{S}(u_n)$ 与 $x_n' \in \mathcal{S}(u_n')$ 收敛到 x_0, 满足 $x_n' - x_n = \alpha_n h_n, \alpha_n \geqslant 0, \|h_n\| = 1, \alpha_n^{-1}(u_n - u_n') \to 0$. 由于

$$\langle F(x_n, u_n), x_n' - x_n\rangle \geqslant 0, \quad \langle F(x_n', u_n'), x_n - x_n'\rangle \geqslant 0, \tag{5.13}$$

则对充分大的 n, $|\langle F(x_0, u_0), h_n\rangle| < \eta$. 同时

$$x_n' - x_n = (x_n' - x_0) + (x_0 - x_n) \in T_K(x_0) - T_K(x_0),$$

因此 $h_n \in \mathrm{Sp}(T_K(x_0))$. 将 (5.13) 中的不等式相加可得

$$Q_0(x_n' - x_n) \leqslant \langle D_u F(x_0, u_0)(u_n - u_n'), x_n' - x_n\rangle + o(\alpha_n^2) \leqslant o(\alpha_n^2).$$

得到 $Q_0(h_n) \leqslant o(1)$, 因此对充分大的 n, 有 $Q_0(h_n) < \beta$, 这与 (5.11) 相矛盾. □

注 5.3　由引理 3.75 可得, 若空间 X 是自反的, 二次型 Q_0 是 Legendre 的, 则充分条件 (5.7) 等价于条件 (5.9). 此种情形, 由引理 3.75 可得, 条件 (5.11) 等价于下述条件

$$Q_0(h) > 0, \quad \forall h \in \mathrm{cl}[\mathrm{Sp}(T_K(x_0))] \cap DF(x_0,u_0)^{\perp}, \quad h \neq 0. \tag{5.14}$$

注 5.4　在 (5.10) 中取 $u = u_0$, 则条件 (5.7) 可推出 x_0 是非扰动问题 (5.1) 的局部唯一解. 由 (5.12)(通过取 $u_1 = u_2$), 强的充分条件 (5.11) 可推出, 当 u 充分接近于 u_0 时, 相应的变分不等式 (VI_u) 在接近 x_0 处至多有一解. 然而, 并没有声明, 充分条件 (5.7) 或 (5.11) 推出这样一个解的存在性.

现在设 K 是多面性的 (见定义 3.51), 对给定的方向 $d \in U$, 讨论沿形式为 $u_t := u_0 + td + o(t)$ 的路径的扰动分析. 考虑下述线性化变分不等式：求 $h \in C(x_0)$ 满足

$$\langle DF(x_0,u_0)(h,d), z-h\rangle \geqslant 0, \quad \forall z \in C(x_0). \tag{5.15}$$

等价地, 上述线性化变分不等式 (5.15) 可以写为下述形式：求 h 满足

$$-DF(x_0,u_0)(h,d) \in N_{C(x_0)}(h). \tag{5.16}$$

若存在实值函数 $f(x,u)$ 满足 $F(x,u) = D_x f(x,u)$, 上述线性化变分不等式表示二次优化问题的最优性系统：

$$\min_{h\in C(x_0)} D^2 f(x_0,u_0)((h,d),(h,d)).$$

定理 5.5　*考虑路径 $u_t : u_0 + td + o(t)$. 设*

(i) 空间 X 是自反的.

(ii) $\mathcal{S}(u_0) = \{x_0\}$.

(iii) 多值函数$t \to \mathcal{S}(u_t), t > 0$ 在 $t = 0$ 处是上 Lipschitz 连续的, 即存在 $c > 0$ 满足对所有 $t > 0$ 充分小满足 $\mathcal{S}(u_t) \subset \{x_0\} + ctB_X$.

(iv) 二次型 Q_0 是 Legendre 形式.

(v) 集合 K 在点 x_0 是多面性的.

则对任何 $x_t \in \mathcal{S}(u_t)$, 当 $t \downarrow 0$ 时, $t^{-1}(x_t - x_0)$ 的任何弱极限点 h_0 是 $t^{-1}(x_t - x_0)$ 的强极限点, h_0 是线性化变分不等式 (5.15) 的解. 进一步, 若 (5.15) 有唯一解 $\bar{h}$, 则对任何 $x_t \in \mathcal{S}(u_t)$, 有 $x_t = x_0 + t\bar{h} + o(t)$.

证明　考虑 $x_t \in \mathcal{S}(u_t)$, 其中 $t \geqslant 0$ 使得 $\mathcal{S}(u_t)$ 是非空的. 注意, 若对所有充分小的 $t > 0$, $\mathcal{S}(u_t)$ 是空集, 则结论是显然成立的. 因此, 可设至少存在序列 $t_n \downarrow 0$, 满足这样的 x_{t_n} 是存在的. 因为由假设 (iii), $\|x_t - x_0\| = O(t)$, 可以表示 $x_t = x_0 + th_t$,

其中, 当 $t\downarrow 0$ 时, h_t 是有界的, 因为 X 是自反的 Banach 空间, 它至少有一弱极限点 h_0. 由于 $h_t\in T_K(x_0)$, $T_K(x_0)$ 是凸的, 因而在弱拓扑下为闭的, 有 $h_0\in T_K(x_0)$. 又因为 $x_t\in\mathcal{S}(u_t)$, 有

$$\langle F(x_t,u_t),y-x_t\rangle\geqslant 0,\quad \forall y\in K. \tag{5.17}$$

结果, 取 $y=x_0$, 可得到 $\langle F(x_t,u_t),h_t\rangle\leqslant 0$, 因此有 $\langle F(x_0,u_0),h_0\rangle\leqslant 0$. 因为由 (5.6), 对所有的 $h\in T_K(x_0)$, $\langle F(x_0,u_0),h_0\rangle\geqslant 0$, 得到 $\langle F(x_0,u_0),h_0\rangle=0$. 因此 $h_0\in C(x_0)$.

以下证明 h_0 是线性化变分不等式 (5.15) 的解. 考虑点 $z\in\mathcal{R}_K(x_0)\cap F(x_0,u_0)^{\perp}$. 注意到, 由于 K 的多面性质, 这样的点构成了 $C(x_0)$ 的稠密子集, 因此只需要对这样的点验证 (5.15) 中的不等式. 因为 $z\in\mathcal{R}_K(x_0)$, 对某一 $\tau>0$, 有 $x_0+\tau z\in K$. 则 $z_t:=z+\tau^{-1}(x_0-x_t)$ 满足 $x_t+\tau z_t=x_0+\tau z\in K$, 且当 $t\downarrow 0$ 时 $z_t\to z$. 在 (5.17) 中取 $y=x_t+\tau z_t$, 得到

$$\langle F(x_t,u_t),z_t\rangle\geqslant 0.$$

由 z_t 的表达式, 由 $z\in C(x_0)$, 有

$$\begin{aligned}0\leqslant&\ \langle F(x_t,u_t),z_t\rangle\\ =&\ \langle F(x_0,u_0)+tDF(x_0,u_0)(h_t,d),z_t\rangle+o(t)\\ =&\ \tau^{-1}\langle F(x_0,u_0),x_0-x_t\rangle+t\langle DF(x_0,u_0)(h_t,d),z\rangle+o(t).\end{aligned}$$

由 (5.6), 上面表达式右端的第一项是非正的, 将上述不等式除以 t, 再取极限, 得到

$$\langle DF(x_0,u_0)(h_0,d),z\rangle\geqslant 0. \tag{5.18}$$

另一方面, 在 (5.17) 中取 $y=x_0$, 得到

$$\langle F(x_0,u_0)+tDF(x_0,u_0)(h_t,d),h_t\rangle\leqslant o(t).$$

因为由 (5.6) 可得到 $\langle F(x_0,u_0),h_t\rangle\geqslant 0$, 于是有

$$\langle DF(x_0,u_0)(h_t,d),h_t\rangle\leqslant o(1). \tag{5.19}$$

因为二次型 Q_0 是 Legendre 的, 因而是弱下半连续的, 可得

$$\langle DF(x_0,u_0)(h_0,d),h_0\rangle\leqslant 0.$$

结合 (5.18), 可推出

$$\langle DF(x_0,u_0)(h_0,d),z-h_0\rangle\geqslant 0,\quad \forall z\in C(x_0)\cap\mathcal{R}_K(x_0),$$

由此, (5.15) 可由 K 的多面性得到.

进一步, 由 (5.15) 可得 $\langle DF(x_0,u_0)(h_0,d),h_0\rangle=0$. 因为 Q_0 是 Legendre 形式, 由 (5.19), 对满足 h_{t_n} 弱收敛于 h_0 的序列 $t_n\downarrow 0$, 可得

$$\begin{aligned}Q_0(h_0)&=\langle D_xF(x_0,u_0)h_0,h_0\rangle=-\langle D_uF(x_0,u_0)d,h_0\rangle\\&=-\lim_{n\to\infty}\langle D_uF(x_0,u_0)d,h_{t_n}\rangle\geqslant\liminf_{n\to\infty}\langle D_xF(x_0,u_0)h_{t_n},h_{t_n}\rangle.\end{aligned}$$

由此得 h_{t_n} 是强收敛到 h_0 的.

最后, 若 $x_t\in\mathcal{S}(u_t)$ 且 (5.15) 具有唯一解 $\bar{h}$, 则 $t^{-1}(x_t-x_0)$ 的任何弱极限点均等于 $\bar{h}$. 因为 X 是自反的, 且对任何 $t_n\downarrow 0$, 序列 $t_n^{-1}(x_{t_n}-x_0)$ 是有界的, 则它有弱极限点, 也是强极限点. 最后的结论得证. □

可以给出注记, 由上述证明可得, 若 $\mathcal{S}(u_{t_n})$ 对某一序列 $t_n\downarrow 0$ 是非空的, 则上述定理的假设 (i)~(v) 可推出线性化变分不等式 (5.15) 解的存在性.

5.1.2 广义方程

这一节考虑下述参数化的广义方程: 求 $(x,\lambda)\in X\times Y^*$ 满足

$$(\mathrm{GE}_u)\qquad F(x,u)+D_xG(x,u)^*\lambda=0,\quad \lambda\in N_K(G(x,u)),\tag{5.20}$$

其中 X,Y 与 U 是 Banach 空间, $K\subset Y$ 是非空的闭凸集, $F:X\times U\to X^*$ 是连续可微映射, $N_K(y)$ 记 K 在 y 处的法锥(若 $y\notin K$, 它被设为空集). 本节设空间 Y 或者是自反的或者是可分的, 从而 Y^* 中的任何有界序列均有弱 * 收敛的子序列.

注 5.6　若 $F(x,u)=D_xf(x,u)$, 其中 $f:X\times U\to\mathbb{R}$, 则 (5.20) 表示 (以 $u\in U$ 为参数的) 在 $G(x,u)\in K$ 约束下的 $f(x,u)$ 极小化最优化问题的最优性系统. 4.4.3 与 4.4.4 节简短讨论了这样的广义方程. 若 $G(x,u):=x$ 是单位映射, 则上述广义方程变为 $\lambda=-F(x,u)$ 且 $\lambda\in N_K(x)$, 因此可以写为形式 $-F(x,u)\in N_K(x)$. 所以, 在此种情况下, 广义方程(5.20) 与参数化变分不等式 (VI_u) 是重合的, 其中 (VI_u) 由 (5.6) 给出.

注 5.7　包含关系 $\lambda\in N_K(G(x,u))$ 可表示为等价的形式 $G(x,u)\in N_K^{-1}(\lambda)$, 因此可将广义方程(5.20) 写为下述形式

$$(F(x,u)+D_xG(x,u)^*\lambda,G(x,u))\in\{0\}\times N_K^{-1}(\lambda).\tag{5.21}$$

尤其, 若集合 K 是闭凸锥, 有

$$N_K^{-1}(\lambda)=N_{K^-}(\lambda)=\begin{cases}K\cap\ker\lambda, & 若\lambda\in K^-,\\ \varnothing, & 否则,\end{cases}\tag{5.22}$$

其中 K^- 是锥 K 的极锥. 因为 $\{0\} \times N_{K^-}(\lambda)$ 与 $X \times K^-$ 在 (x, λ) 处的法锥重合, 可视 (5.21) 为关于集合 $X \times K^-$ 与映射

$$\varphi_u(x, \lambda) := (F(x, u) + D_x G(x, u)^* \lambda, -G(x, u)) \tag{5.23}$$

的参数变分不等式, 其中 φ_u 是 Banach 空间$X \times Y^*$ 到其对偶 $X^* \times Y$ 的映射.

注 5.8 由某一约束规范可得

$$N_{\Phi(u)} = D_x G(x, u)^* [N_K(G(x, u))],$$

其中 $\Phi(u) := \{x \in X : G(x, u) \in K\}$. 因此, 在某一约束规范下, 可以把广义方程(5.20) 写为形式

$$-F(x, u) \in N_{\Phi(u)}(x). \tag{5.24}$$

进一步, 若映射 G 不依赖于 u, 则集合 $\Phi := G^{-1}(K)$ 是与 u 无关的, 广义方程(5.20) 变成相对于集合 Φ(可能是非凸的) 的参数变分不等式 $-F(x, u) \in N_\Phi(x)$.

用 $\mathcal{S}(u)$ 记广义方程(5.20) 的解集. 本节假定非扰动广义方程有唯一解 (x_0, λ_0), 即 $\mathcal{S}(u_0) = \{(x_0, \lambda_0)\}$. 考虑集合

$$K_0 := \{y \in K : \langle \lambda_0, y - y_0 \rangle = 0\}, \quad \mathcal{K} := \{y \in T_K(y_0) : \langle \lambda_0, y \rangle = 0\},$$

其中 $y_0 := G(x_0, u_0)$. 显然 $\mathcal{K}$ 是凸锥, $y_0 \in K_0$ 且 $T_{K_0}(y_0) \subset \mathcal{K}$. 进一步, 若集合 K 在 y_0 处为多面的, 则 $T_{K_0}(y_0) = \mathcal{K}$. 若 Robinson 约束规范关于简约集 K_0 是成立的, 则严格约束规范在 x_0 处相对于 λ_0 成立 (见定义 4.46). 严格约束规范意味着乘子 λ_0 的唯一性 (见命题 4.47).

非扰动广义方程在解 x_0 处的临界锥定义为

$$\begin{aligned} C(x_0) &:= \{h \in X : DG(x_0)h \in T_K(y_0), \langle F(x_0), h \rangle = 0\} \\ &= \{h \in X : DG(x_0)h \in \mathcal{K}\}, \end{aligned}$$

其中 $F(\cdot) = F(\cdot, u_0)$ 与 $G(\cdot) = G(\cdot, u_0)$ 是相应的非扰动映射, $y_0 := G(x_0)$. 类似地, 可以定义近似临界锥

$$\widehat{C}_\eta(x_0) := \{h \in X : DG(x_0)h \in T_K(G(x_0)), \langle F(x_0), h \rangle \leqslant \eta \|h\|\}.$$

最后, 与 (x_0, λ_0) 相联系的二次型定义为

$$Q_0(h) := \langle DF(x_0)h, h \rangle + \langle \lambda_0, D^2 G(x_0)(h, h) \rangle. \tag{5.25}$$

应该注意到, 即使 K 是锥, 也不能以一种直接的方式将命题 5.2 中的稳定性结果用于变分不等式 (5.21), 因为对应于 (5.7) 与 (5.8) 的充分性条件对这样的变分不

等式不成立. 然而, 可将定理 4.51 中相同的推证用于广义方程的框架. 考虑 (4.116) 中定义的多值映射 Ω, 即

$$\Omega(a,y) := \{\lambda \in Y^* : [DG(x_0)]^*\lambda + a = 0, \lambda \in N_K(y)\}. \tag{5.26}$$

注意到, 对 $a_0 = F(x_0), y_0 = G(x_0)$, 集合 $\Omega(a_0, y_0)$ 与满足 $(x_0, \lambda) \in \mathcal{S}(u_0)$ 的 λ 的集合是重合的, 严格约束规范可推出 $\Omega(a_0, y_0) = \{\lambda_0\}$, 且多值映射 Ω 在 (a_0, y_0) 处是上 Lipschitz 连续的. 下述定理的证明类似于定理 4.51 的证明.

定理 5.9 令 $(\bar{x}(u), \overline{\lambda}(u)) \in \mathcal{S}(u)$ 是 (GE_u) 的解. 设

(i) 当 $u \to u_0$ 时, $\bar{x}(u)$ 收敛到 x_0.

(ii) 在 x_0 处相对于 λ_0 的严格约束规范成立.

(iii) 下述二阶条件成立: 存在常数 $\beta > 0$ 与 $\eta > 0$ 满足

$$Q_0(h) \geqslant \beta \|h\|^2, \quad \forall h \in \widehat{C}_\eta(x_0). \tag{5.27}$$

则

$$\|\bar{x}(u) - x_0\| + \|\overline{\lambda}(u) - \lambda_0\| = O(\|u - u_0\|). \tag{5.28}$$

对给定的方向 $d \in U$, 考虑集合

$$\begin{aligned} \mathcal{C}_d(x_0) &:= \{h \in X : DG(x_0, u_0)(h, d) \in \mathcal{K}\} \\ &= \{h \in X : DG(x_0, u_0)(h, d) \in T_K(y_0) \cap \{\lambda_0\}^\perp\}. \end{aligned}$$

若 $F(x,u) = D_x f(x,u)$, 因而 λ_0 是相应的最优化问题的 (唯一的)Lagrange 乘子, 则 $\mathcal{C}_d(x_0)$ 与线性化最优化问题 (PL_d) 的最优解集相同, 其中 (PL_d) 由 (4.45) 定义. 对于 $d = 0$, 这一集合变为临界锥$C(x_0)$.

考虑在点 (x_0, λ_0) 处沿方向 $d \in U$ 的线性化广义方程: 求 $(h, \mu) \in X \times Y^*$ 满足

$$\begin{aligned} &\text{(i) } D_{(x,u)}[F(x_0,u_0) + D_x G(x_0,u_0)^*\lambda_0](h,d) + D_x G(x_0,u_0)^*\mu = 0. \\ &\text{(ii) } DG(x_0,u_0)(h,d) \in N_{\mathcal{K}^-}(\mu). \end{aligned} \tag{5.29}$$

注意到, 上述条件 (5.29)(ii) 等价于

$$DG(x_0,u_0)(h,d) \in \mathcal{K}, \quad \mu \in \mathcal{K}^- \text{ 且 } \langle \mu, DG(x_0,u_0)(h,d) \rangle = 0. \tag{5.30}$$

若 $G(x,u) := x$ 是单位映射, 则广义方程(5.29) 简化为线性化变分不等式 (5.15). 若 $F(x,u) = D_x f(x,u)$, 则广义方程(5.29) 表示的是下述二次规划问题的最优性系统

$$\min_{h \in \mathcal{C}_d(x_0)} D^2 L(x_0, \lambda_0, u_0)((h,d),(h,d)),$$

其中 $L(x,\lambda,u) := f(x,u) + \langle \lambda, G(x,u)\rangle$ 是相应的 Lagrange 函数. 下述定理是定理 5.5 的一个推广.

定理 5.10 考虑形式为 $u_t := u_0 + td + o(t^2)$ 形式的路径. 设

(i) 空间 X 是自反的.

(ii) $\mathcal{S}(u_0) = \{(x_0, \lambda_0)\}$.

(iii) 多值映射 $t \to \mathcal{S}(u_t), t \geqslant 0$ 在 $t = 0$ 处是上 Lipschitz 连续的.

(iv) 由 (5.25) 定义的二次型 Q_0 是 Legendre 形式.

(v) 集合 K 在 $y_0 := G(x_0, u_0)$ 处是多面性的.

(vi) 约束 $G(\cdot, u_0) \in K$ 在点 x_0 处的 Robinson 约束规范成立.

则当 $t \downarrow 0$ 时, 对任何 $(x_t, \lambda_t) \in \mathcal{S}(u_t)$ 及 $t^{-1}(x_t - x_0, \lambda_t - \lambda_0)$ 的任何弱极限点 (h_0, μ_0), h_0 是 $t^{-1}(x_t - x_0)$ 的强极限点, 且 (h_0, μ_0) 是线性化广义方程 (5.29) 的解. 进一步, 若 (5.29) 具有唯一解 $(\bar{h}, \bar{\mu})$, 则对任何 $(x_t, \lambda_t) \in \mathcal{S}(u_t)$, 当 $t \downarrow 0$ 时, $x_t = x_0 + t\bar{h} + o(t)$ 且 $t^{-1}(\lambda_t - \lambda_0)$ 分别以强拓扑与弱拓扑收敛到 $\bar{\mu}$.

证明 若对所有充分小的 $t \geqslant 0$, $\mathcal{S}(u_t)$ 均是空集, 则结论显然成立. 因此, 设至少存在序列 $t_n \downarrow 0$, 使得 $\mathcal{S}(u_{t_n})$ 是非空集.

令 $(x_t, \lambda_t) \in \mathcal{S}(u_t)$. 因为 $\mathcal{S}(u_t)$ 是上方 Lipschitz 连续的, 可表示 $x_t = x_0 + th_t$ 与 $\lambda_t = \lambda_0 + t\mu_t$, 其中当 $t \downarrow 0$ 时, (h_t, μ_t) 是有界的. 因为 X 是自反的, (h_t, μ_t) 至少有一弱极限点 $(h_0, \mu_0) \in X \times Y^*$. 由方程

$$F(x_t, u_t) + D_x G(x_t, u_t)^* \lambda_t = 0$$

在点 (x_0, u_0) 处的一阶展式得到

$$D_{(x,u)}[F(x_0, u_0) + D_x G(x_0, u_0)^* \lambda_0](h_t, d) + D_x G(x_0, u_0)^* \mu_t = o(1). \tag{5.31}$$

于是得到 (h_0, μ_0) 满足方程 (5.29)(i).

因为 $G(x_0 + th_t, u_t) \in K$, 有 $DG(x_0, u_0)(h_0, d) \in T_K(y_0)$. 下证

$$\langle \lambda_0, DG(x_0, u_0)(h_0, d)\rangle = 0. \tag{5.32}$$

因为 $G(x_0, u_0) \in K$ 且 $\lambda_t \in N_K(G(x_t, u_t))$, 有

$$\langle \lambda_t, G(x_0, u_0) - G(x_t, u_t)\rangle \leqslant 0,$$

因此有 $\langle \lambda_0, DG(x_0, u_0)(h_0, d)\rangle \geqslant 0$. 因为由 $\lambda_0 \in N_K(y_0)$ 可得相反的不等式, 方程 (5.32) 得证. 等式 (5.32) 与 $DG(x_0, u_0)(h_0, d) \in T_K(y_0)$ 可推出 $DG(x_0, u_0)(h_0, d) \in \mathcal{K}$.

因为 (5.30) 中的第二与第三条件等价于条件 $\mu_0 \in N_{\mathcal{K}}(DG(x_0, u_0)(h_0, d))$, 为证得 (5.29)(ii), 只需证明

$$\langle \mu_0, z - DG(x_0, u_0)(h_0, d)\rangle \leqslant 0, \quad \forall z \in \mathcal{K}. \tag{5.33}$$

由 K 的多面性, 只需证明当 $z \in \mathcal{R}_K(y_0) \cap (\lambda_0)^{\perp}$ 时上述不等式成立. 令 $\bar{t} > 0$ 满足 $y_0 + \bar{t}z \in K$, 则对所有 $t \in (0, \bar{t})$, 有

$$\langle \lambda_0 + t\mu_t, y_0 + tz - G(x_t, u_t) \rangle \leqslant 0, \quad \langle \lambda_0, G(x_t, u_t) - y_0 \rangle \leqslant 0.$$

将这些不等式相加并由 $\langle \lambda_0, z \rangle = 0$, 得到

$$\langle \mu_t, z - DG(x_0, u_0)(h_t, d) \rangle \leqslant o(1). \tag{5.34}$$

另一方面, 考虑二次函数

$$Q_d(h) := \langle D_{(x,u)}[F(x_0, u_0) + D_x G(x_0, u_0)^* \lambda_0](h, d), (h, d) \rangle. \tag{5.35}$$

考虑序列 $t_n \downarrow 0$, 它满足 (h_{t_n}, μ_{t_n}) 弱收敛到 (h_0, μ_0). 因为 Q_0 是 Legendre 形式, 所以 Q_d 是弱下半连续的, 且若 $h_{t_n} \xrightarrow{w} h_0$ 且 $Q_d(h_{t_n}) \to Q_d(h_0)$, 则 $h_{t_n} \to h_0$. 由 (5.29)(i) 与 (5.31) 得到

$$\begin{aligned} -\langle \mu_0, D_x G(x_0, u_0)(h_0, d) \rangle = Q_d(h_0) &\leqslant \liminf_{n \to \infty} Q_d(h_{t_n}) \\ &= \liminf_{n \to \infty} -\langle \mu_{t_n}, D_x G(x_0, u_0)(h_{t_n}, d) \rangle. \end{aligned}$$

将此与 (5.34) 相结合, 得到 (5.33). 于是证得 (h_0, μ_0) 是线性化的广义方程 (5.29) 的解.

因为 $\langle \mu_0, D_x G(x_0, u_0)(h_0, d) \rangle = 0$, 得到 $Q_d(h_0) = 0$. 进一步, 在 (5.34) 中取 $z = 0$, 得到

$$-\langle \mu_{t_n}, D_x G(x_0, u_0)(h_{t_n}, d) \rangle \leqslant o(t_n),$$

因此有

$$0 = Q_d(h_0) \leqslant \liminf_{n \to \infty} Q_d(h_{t_n}) \leqslant 0.$$

可推出 $h_{t_n} \to h_0$. 最后, 若线性化广义方程(5.29) 有唯一解 $(\bar{h}, \bar{\mu})$, 则最后结论得证. □

注 5.11 若 $K \subset Y$ 是闭凸锥, 则 K 在点 $y_0 := 0 \in Y$ 处总是多面性的, 这是因为 $\mathcal{R}_K(0) = T_K(0) = K$. 所以, 若 K 是闭凸锥, $G(x_0, u_0) = 0$, 则上述定理中的假设 (v) 是自动成立的. 在某些情况下, 集合 K 可以通过适当的变换 Ξ 简化为锥, 其中 $\Xi(y_0) = 0$(见定义 3.135). 这种情况发生在例如 K 是正半定矩阵锥时 (见例 3.140). 在这种情况下, 上述定理的结论可应用于相应的简化问题.

5.1.3 强正则性

这一节讨论隐函数定理(见定理 5.14) 中用到的正则条件的一个推广. 基本思想是将非线性方程的适定性简化为相应的线性化系统的适定性. 通过对由映射 $F(x, u)$

与 $G(x,u)$ 的线性化得到的广义方程实施扰动分析, 这一思想被推广到 (5.20) 型的广义方程. 注意到集合 K 本身在任何意义下均不是线性化的. 一方面, 这允许我们研究扰动的广义方程的适定性并且推导各种各样的稳定性估计. 另一方面, $F(x,u)$ 与 $G(x,u)$ 本身的线性化对计算解的方向导数是不够的, 因为这样的线性化系统关于所考虑的变量不是正齐次的, 即计算解的方向导数也需要集合 K 的线性化. 在某些具体情况下, 这在定理 5.5 与定理 5.10 中实现过.

通过考虑下述 "抽象" 广义方程框架下的部分线性化开始我们的讨论: 求 $z \in Z$ 满足

$$0 \in \varphi(z) + N(z), \tag{5.36}$$

其中 Z 与 W 是 Banach 空间, $\varphi: Z \to W$ 是连续可微的映射, $N: Z \to 2^W$ 是多值映射. 可以视上述的广义方程与其参数化 (5.38) 为广义方程 (5.21) 的抽象形式, 其中 $Z := X \times Y^*, W := X^* \times Y, N(x,\lambda) := \{0\} \times N_K^{-1}(\lambda)$, φ 由 (5.23) 给出.

设 z_0 是 (5.36) 的解, 与下述 (部分的) 线性化抽象广义方程相联系: 求 $z \in Z$ 满足

$$(\mathrm{LE}_\delta) \qquad \delta \in \varphi(z_0) + D\varphi(z_0)(z - z_0) + N(z), \tag{5.37}$$

其中 $\delta \in W$ 是参数.

定义 5.12 称 z_0 是抽象广义方程(5.36) 的强正则解, 若存在 $z_0 \in Z$ 与 $0 \in W$ 的邻域 $\mathcal{V}_Z$ 与 $\mathcal{V}_W$, 满足对每一 $\delta \in \mathcal{V}_W$, 线性化广义方程(LE_δ) 在 $\mathcal{V}_Z$ 中有唯一解, 记为 $\varsigma(\delta)$, 映射 $\varsigma: \mathcal{V}_W \to \mathcal{V}_Z$ 是常数为 c 的 Lipschitz 连续的映射.

在上述定义中, 不排除邻域 $\mathcal{V}_Z$ 之外 (LE_δ) 的其他解的可能性. 类似的注记适用于本节其他命题.

令 U 是 Banach 空间, $u_0 \in U$, 令 $\Phi: Z \times U \to W$ 是连续可微的映射, 满足 $\Phi(\cdot, u_0) = \varphi(\cdot)$. 下述定理表明, 解 $\varsigma(\cdot)$ 的上述 Lipschitz 连续性质可以被具有下述形式的广义方程的非线性扰动保持:

$$0 \in \Phi(z,u) + N(z). \tag{5.38}$$

进一步, 扰动解 "接近于" 相应的线性化的广义方程的解: 求 $z \in Z$ 满足

$$0 \in \Phi(z_0,u_0) + D\Phi(z_0,u_0)(z - z_0, u - u_0) + N(z). \tag{5.39}$$

注意到, 对 $\delta_u := -D_u\Phi(z_0,u_0)(u-u_0)$, 且对所有充分接近于 u_0 的 u, $\varsigma(\delta_u)$ 是 (5.39) 在邻域 $\mathcal{V}_Z$ 中的解.

定理 5.13 令 z_0 是抽象广义方程(5.36) 的强正则解, $\varsigma(\delta)$ 是相应的线性化广义方程(LE_δ) 的解, 则分别存在 z_0 与 u_0 的邻域 $\mathcal{V}_Z$ 与 $\mathcal{V}_U$, 满足对每一 $u \in \mathcal{V}_U$, 扰

动的抽象广义方程(5.38) 有一个且仅有一个解 $z(u) \in \mathcal{V}_Z$, 且存在某一 $\kappa > 0$, 对所有的 $u, u' \in \mathcal{V}_U$ 满足

$$\|z(u') - z(u)\| \leqslant \kappa \|u' - u\| \tag{5.40}$$

且

$$z(u) = \varsigma(-D_u\Phi(x_0, u_0)(u - u_0)) + o(\|u - u_0\|). \tag{5.41}$$

证明　用不动点定理证明解的存在性与 (5.40). 令

$$r(z, u) := \varphi(z_0) + D\varphi(z_0)(z - z_0) - \Phi(z, u).$$

包含关系 (5.38) 可表示为

$$r(z, u) \in \varphi(z_0) + D\varphi(z_0)(z - z_0) + N(z). \tag{5.42}$$

因为 $r(z_0, u_0) = \varphi(z_0) - \Phi(z_0, u_0) = 0$, 对充分接近于 (z_0, u_0) 的所有 (z, u), 余项 $r(z, u)$ 属于映射 ς 所定义的邻域 $\mathcal{V}_W$, 因此 $T_u(z) := \varsigma(r(z, u))$ 是有定义的. 包含关系 (5.42) 等价于方程 $z = T_u(z)$.

令 c 是 $\varsigma(\cdot)$ 在 $\mathcal{V}_W$ 上的 Lipschitz 常数, α 是 Φ 在 $\mathcal{V}_Z \times \mathcal{V}_U$ 上的 Lipschitz 常数, $\varepsilon \in \left(0, \frac{1}{3}c^{-1}\right)$. 令 $z \in B(z_0, \rho_1), u \in B(u_0, \rho_2)$, 其中 ρ_1 与 ρ_2 是正的常数, 则对充分小的 $\rho_1 > 0$ 与 $\rho_2 > 0$, 下述不等式成立:

$$\begin{aligned}\|r(z, u)\| &\leqslant \|\Phi(z_0, u_0) + D_z\Phi(z_0, u_0)(z - z_0) - \Phi(z, u_0)\| + \|\Phi(z, u_0) - \Phi(z, u)\| \\ &\leqslant \varepsilon\|z - z_0\| + \alpha\|u - u_0\| \leqslant \varepsilon\rho_1 + \alpha\rho_2.\end{aligned}$$

如有必要, 减小 ρ_1 与 ρ_2, 对在闭球$\overline{B}(z_0, \rho_1)$ 中所有的 z, 余项 $r(z, u)$ 属于 $\mathcal{V}_W$, 因此由强正则性得

$$\|T_u(z) - z_0\| = \|\varsigma(r(z, u)) - z_0\| \leqslant c(\varepsilon\rho_1 + \alpha\rho_2) \leqslant \frac{1}{3}\rho_1 + c\alpha\rho_2 < \rho_1,$$

其中, 最后一个不等式, 当 $\rho_2 < \rho_1/(3c\alpha)$ 时成立. 于是, 对任何 $u \in B(u_0, \rho_2)$, T_u 映 $\overline{B}(z_0, \rho_1)$ 到其自身.

下面证明 T_u 是压缩的. 因为 Φ 是连续可微, 如果有必要可以减小 ρ_1 与 ρ_2, 对于 $z_1, z_2 \in B(z_0, \rho_1)$, $u \in B(u_0, \rho_2)$, 有

$$\begin{aligned}\|r(z_2, u) - r(z_1, u)\| &\leqslant \|\Phi(z_2, u) - \Phi(z_1, u) - D_z\Phi(z_0, u_0)(z_2 - z_1)\| \\ &\leqslant \varepsilon\|z_2 - z_1\|.\end{aligned}$$

由强正则性条件可以得到

$$\|T_u(z_2) - T_u(z_1)\| \leqslant c\|r(z_2, u) - r(z_1, u)\| \leqslant c\varepsilon\|z_2 - z_1\| \leqslant \frac{1}{3}\|z_2 - z_1\|,$$

因此 T_u 是压缩的. 因为 T_u 在完备度量空间$\overline{B}(z_0,\rho_1)$ 内是压缩的, 由不动点定理(定理 2.2), 存在不动点 $\bar{z}\in\overline{B}(z_0,\rho_1)$. 得到 $\bar{z}$ 是 (5.38) 的解.

设 $z_1=z(u_1), z_2=z(u_2)$ 是在 $\overline{B}(z_0,\rho_1)$ 中 (5.38) 的解, 对应着向量 $u_1,u_2\in B(u_0,\rho_2)$. 得到

$$r(z_i,u_i)\in\varphi(z_0)+D\varphi(z_0)(z_i-z_0)+N(z_i),\quad i=1,2,$$

因此

$$\begin{aligned}\|r(z_2,u_2)-r(z_1,u_1)\|&\leqslant\|r(z_2,u_2)-r(z_1,u_2)\|+\|r(z_1,u_2)-r(z_1,u_1)\|\\&\leqslant\varepsilon\|z_2-z_1\|+\alpha\|u_2-u_1\|.\end{aligned}$$

所以, 由强正则性,

$$\|z_2-z_1\|\leqslant c\|r(z_2,u_2)-r(z_1,u_1)\|\leqslant\frac{1}{2}\|z_2-z_1\|+c\alpha\|u_2-u_1\|,$$

可推出 $z(u)$ 在 $B(u_0,\rho_2)$ 上是以 $2c\alpha$ 为模的 Lipschitz 连续的. 于是得到 $z(u)$ 的唯一性.

最后, 对 $u\in B(u_0,\rho_2)$, 有 $z(u)$ 是 T_u 的不动点, 因此

$$z(u)=\varsigma(r(z(u),u))=\varsigma(-D_u\Phi(x_0,u_0)(u-u_0)+o(\|u-u_0\|)),$$

由此, 由 $\varsigma(\cdot)$ 的 Lipschitz 连续性可得到 (5.41). □

在某种意义下, 上述定理可视为如下表述的经典隐函数定理的推广.

定理 5.14 *设映射 $\Phi(z,u)$ 是连续可微的, $\Phi(z_0,u_0)=0$, 线性映射 $D_z\Phi(z_0,u_0):Z\to W$ 是一对一的且映上的, 则对接近于 u_0 的所有 u, 方程 $\Phi(z,u)=0$ 在 z_0 附近有唯一解, 在 u_0 的邻域内是 Lipschitz 连续的且 Fréchet 可微的, 且*

$$Dz(u_0)=-D_z\Phi(z_0,u_0)^{-1}[D_u\Phi(z_0,u_0)].\tag{5.43}$$

证明 置多值映射 $N(z)$ 恒等于零. 广义方程 (5.38) 简化为方程 $\Phi(z,u)=0$, 线性化系统 (5.37) 简化为 $D_x\Phi(x_0,u_0)(z-z_0)=\delta$. z_0 是强正则的充分必要条件是 $D_x\Phi(z_0,u_0):Z\to W$ 是一对一的且映上的, 相应的常数 c 等于 $\|D_x\Phi(z_0,u_0)^{-1}\|$, 由开映射定理, 它是有限的. 于是由定理 5.13 中的相应的结果得到结论. □

由定理 5.13, 强正则性条件在小的扰动下是稳定的, 也可将定理 5.13 的结果以非参数化形式描述. 设 V 是 z_0 的开邻域, 考虑由连续可微映射 $\psi:V\to W$, 采用有限范数

$$\|\psi\|_1:=\sup_{z\in V}\|\psi(z)\|+\sup_{z\in V}\|D\psi(z)\|$$

构成的 Banach 空间$U := C^1(V, W)$. 定义由 $V \times U$ 到 W 的映射 $\Phi(z, u) := u(z)$, 令 $u_0 := \varphi$. 这一映射 Φ 在 $V \times U$ 上是连续可微的, 因此, 由定理 5.13, 有下述结论: 若 z_0 是抽象广义方程(5.36) 的强正则解, 则对 φ 邻域中所有的 u (关于范数$\|\cdot\|_1$), 广义方程 $0 \in u(z) + N(z)$ 在 z_0 的邻域内有 Lipschitz 连续的 (因而是唯一的) 解 $\bar{z}(u)$, 且

$$\bar{z}(u) = \varsigma(\varphi(z_0) - u(z_0)) + o(\|u - \varphi\|_1). \tag{5.44}$$

下述结论是定理 5.13 的结果.

推论 5.15　*设 z_0 是抽象广义方程(5.36) 的强正则解. 则分别存在 $z_0, 0 \in W$ 与 $D\varphi(z_0) \in \mathcal{L}(Z, W)$ 的邻域 $\mathcal{V}_Z, \mathcal{V}_W$ 与 $\mathcal{V}_L$, 满足对每一 $(\delta, A) \in \mathcal{V}_W \times \mathcal{V}_L$, 广义方程*

$$\delta \in \varphi(z_0) + A(z - z_0) + N(z) \tag{5.45}$$

具有唯一解 $z(\delta, A) \in \mathcal{V}_Z$. 进一步, 存在 $\kappa > 0$ 满足, 对任何 $(\delta, A), (\delta', A') \in \mathcal{V}_W \times \mathcal{V}_L$, 下述不等式成立

$$\|z(\delta', A') - z(\delta, A)\| \leqslant \kappa(\|\delta - \delta'\| + \|A' - A\|). \tag{5.46}$$

证明　取 $u := (\delta, A) \in W \times \mathcal{L}(Z, W), u_0$ 等于 $(0, D\varphi(z_0))$ 与 $\Phi(z, u) := \varphi(z_0) - \delta + A(z - z_0)$, 应用定理 5.13 即可. □

一般来说, 很难给出强正则性的充分必要条件, 只能得到部分结果. 这一节的余下部分关注具有 (5.20) 形式且 $F(x)$ 由函数 $f : X \to \mathbb{R}$ 的导数给出的方程, 即 $F(x) = Df(x)$ 的 (非扰动) 广义方程. 考虑下述广义方程: 求 $(x, \lambda) \in X \times Y^*$ 满足

$$0 \in \begin{bmatrix} Df(x) + DG(x)^*\lambda \\ -G(x) \end{bmatrix} + \begin{bmatrix} \{0\} \\ N_K^{-1}(\lambda) \end{bmatrix}. \tag{5.47}$$

假设 $f(x)$ 与 $G(x)$ 是二次连续可微的, 因此映射 $F(x) = Df(x)$ 是连续可微的.

因为条件 $G(x) \in N_K^{-1}(\lambda)$ 等价于 $\lambda \in N_K(G(x))$, 上述广义方程可表示下述最优化问题的一阶最优性条件

$$\min_{x \in X} f(x), \quad \text{s. t.} \quad G(x) \in K. \tag{5.48}$$

即, 若 (x, λ) 是 (5.47) 的解, 则 x 是问题 (5.48) 的稳定点, λ 是相应的 Lagrange 乘子. 称 $(x, \lambda) \in X \times Y^*$ 是问题 (5.48) 的临界点, 如果 (x, λ) 是广义方程 (5.47) 的解, 即 (x, λ) 满足相应的一阶最优性条件.

注意到, 对应于 (5.47), 在 (x_0, λ_0) 处的线性化广义方程 (LE_δ) 可以写为形式: 求 $(h, \mu) \in X \times Y^*$ 满足

$$
\begin{aligned}
&D^2_{xx}L(x_0, \lambda_0)h + DG(x_0)^*\mu = \delta_1,\\
&G(x_0) + DG(x_0)h + \delta_2 \in N_K^{-1}(\lambda_0 + \mu),
\end{aligned}
\tag{5.49}
$$

其中 $L(x, \lambda) := f(x) + \langle \lambda, G(x)\rangle, \delta := (\delta_1, \delta_2) \in X^* \times Y$. 上述线性化广义方程可表示下述问题的一阶最优性条件

$$
\begin{aligned}
\min_{h\in X} \quad & \frac{1}{2}D^2_{xx}L(x_0, \lambda_0)(h, h) - \langle \delta_1, h\rangle,\\
\text{s. t.} \quad & G(x_0) + DG(x_0)h + \delta_2 \in K.
\end{aligned}
\tag{5.50}
$$

若 (x_0, λ_0) 是 (5.47) 的强正则解, 则对充分接近于 $0 \in Y$ 的所有 δ_2, 问题 (5.50) 有可行解, 因此

$$
0 \in \operatorname{int}\{G(x_0) + DG(x_0)X - K\}. \tag{5.51}
$$

这说明强正则性可推出 Robinson 约束规范.

下面研究 x_0 是问题 (5.48) 的局部最优解的情况. 现在引入相对于任意扰动的二阶增长条件的一致形式. 令 U 是 Banach 空间, $f : X \times U \to \mathbb{R}, G : X \times U \to Y$. 称 $(f(x, u), G(x, u)), u \in U$ 是最优化问题 (5.48) 的参数化, 若存在 $u_0 \in U$ 满足 $f(\cdot, u_0) = f(\cdot), G(\cdot, u_0) = G(\cdot)$. 若 $f(x, u)$ 与 $G(x, u)$ 是二次连续可微的, 称这一参数化是 C^2 光滑的. 用 (P_u) 记约束为 $G(\cdot, u) \in K$ 极小化 $f(\cdot, u)$ 的参数化问题.

称参数化是标准的, 若 $U := X^* \times Y, u := (\delta_1, \delta_2)$, 其中 $\delta_1 \in X^*, \delta_2 \in Y, u_0 := (0, 0) \in X^* \times Y$, 且

$$
(f(x, u), G(x, u)) := (f(x) - \langle \delta_1, x\rangle, G(x) + \delta_2). \tag{5.52}
$$

称参数化是倾斜的, 若 $U := X^*, u := \delta_1, u_0 := 0 \in X^*$, 于是有

$$
(f(x, u), G(x, u)) := (f(x) - \langle \delta_1, x\rangle, G(x)). \tag{5.53}
$$

称参数化 $(f(x, u), G(x, u)), u \in U$ 包含标准的参数化, 若 U 可以表示为形式 $U := U' \times X^* \times Y$, 且对 $u = (u', \delta_1, \delta_2) \in U$, 有 $f(x, u) = f(x, u') - \langle \delta_1, x\rangle, G(x, u) = G(x, u') + \delta_2$ 与 $u_0 := (u'_0, 0, 0) \in U$. 倾斜参数化的包含可类似定义. 很显然, 标准参数化包含倾斜参数化.

定义 5.16 令 x_0 是最优化问题 (5.48) 的稳定点. 称一致二阶增长条件在 x_0 处相对于 C^2 光滑参数化 $(f(x, u), G(x, u))$ 成立, 若存在 $c > 0$ 与 x_0 及 u_0 的邻域

$\mathcal{V}_X$ 与 $\mathcal{V}_U$, 满足对任何 $u \in \mathcal{V}_U$ 及相应的参数化问题 (P_u) 的任何稳定点 $\bar{x}(u) \in \mathcal{V}_X$, 下述不等式成立

$$f(x,u) \geqslant f(\bar{x}(u),u) + c\|x - \bar{x}(u)\|^2, \quad \forall x \in \mathcal{V}_X, \quad G(x,u) \in K. \tag{5.54}$$

若对每一 C^2 光滑参数化均是成立的, 称一致二阶增长条件在 x_0 处成立.

标准参数化问题的线性化导致广义方程 (5.49). 对标准参数化, 条件 (5.54) 取下述形式

$$f(x) - \langle \delta_1, x - \bar{x} \rangle \geqslant f(\bar{x}) + c\|x - \bar{x}\|^2, \quad \forall x \in \mathcal{V}_X \cap G^{-1}(K - \delta_2), \tag{5.55}$$

其中 $\bar{x} := \bar{x}(u)$.

如第 4 章看到的那样, (P_u) 的最优解的逐点稳定性 (上 Lipschitz 连续性) 与二阶增长条件是密切相关的. 所以毫不奇怪的是, (P_u) 的稳定点与临界点的定量 (尤其 Lipschitz) 连续性与一致二阶增长条件是相联系的. 通过证明一致二阶增长条件可推出微小扰动的最优化问题局部最优解的存在性.

定理 5.17 设 x_0 是稳定点, $(f(x,u), G(x,u))$ 是最优化问题 (5.48) 的一个 C^2 光滑参数化. 假设所考虑的参数化包含倾斜参数化, (相对于所考虑的参数化的) 一致二阶增长条件在 x_0 点处成立, Robinson 约束规范在 (x_0, u_0) 处成立, 则存在 x_0 与 u_0 的邻域 $\mathcal{V}_X$ 与 $\mathcal{V}_U$ 以及常数 $\kappa > 0$ 满足: 对任何 $u \in \mathcal{V}_U$, 参数化问题 (P_u) 有局部最优解 $\bar{x}(u) \in \mathcal{V}_X$, 且

$$\|\bar{x}(u) - \bar{x}(u')\| \leqslant \kappa \|u - u'\|^{1/2}, \quad \forall u, u' \in \mathcal{V}_U. \tag{5.56}$$

证明 设 $\mathcal{V}_X$ 与 $\mathcal{V}_U$ 是 x_0 与 u_0 的对应于一致二阶增长条件的邻域. 可假设 $\mathcal{V}_X$ 是闭的. 令 $\beta > 0$ 满足 $B(u_0, 2\beta) \subset \mathcal{V}_U$. 给定 $u \in B(u_0, \beta)$, 取 ε_k 是收敛到 0 的正数序列, $x_k, k = 1, 2, \cdots$ 是 $f(\cdot, u)$ 在集合

$$\Psi(u) := \{x \in \mathcal{V}_X : G(x,u) \in K\}$$

上极小化问题的 ε_k^2 最优解. 因为 Robinson 约束规范在小的扰动下是稳定的 (见定理 2.87 后面的注记), 如有必要, 可缩减邻域 $\mathcal{V}_X$ 与 $\mathcal{V}_U$, 可以设映射 $G(\cdot, u)$ 在集合 $\Psi(u)$ 内的每一点 x 处均有 Robinson 约束规范成立, 当然也假设 $\Psi(u)$ 非空.

由 Ekeland 变分原理 (定理 3.23), 存在 $x_k' \in X$, $\delta_k \in X^*$, 满足 x_k' 是 $f(\cdot, u)$ 在 $\Psi(u)$ 上的 ε_k^2 最优解, $\|x_k - x_k'\| \leqslant \varepsilon_k, \|\delta_k\| \leqslant \varepsilon_k \leqslant \beta$ 对充分大的 k 成立, 且 x_k' 是极小化 $\psi_k(\cdot) := f(\cdot, u) - \langle \delta_k, \cdot \rangle$ 于集合 $\Psi(u)$ 上的问题的稳定点. 因为 $u \in B(u_0, \beta)$, 且 $B(u_0, 2\beta) \subset \mathcal{V}_U$, 一致二阶增长条件适用于这一问题. 由命题 4.32 的界, 对任何 $k, m \in \mathbb{N}$, 有 $\|x_k' - x_m'\| \leqslant c^{-1}\|\delta_k - \delta_m\|$. 这表明 x_k' 是 Cauchy 列, 因而收敛到某一点 $\bar{x}(u)$. 于是得到 $\bar{x}(u)$ 是 $f(\cdot, u)$ 在 $\Psi(u)$ 上的局部极小点.

根据命题 4.37 的界, 由一致二阶增长条件可得 Hölder 连续性性质 (5.56), 因为由稳定性定理, 对 $u, u' \in \mathcal{V}_U$, 集合 $\Psi(u)$ 与 $\Psi(u')$ 间的 Hausdorff 距离是 $O(\|u-u'\|)$ 阶的.

注意到对充分小的 β, $\bar{x}(u)$ 不属于 $\mathcal{V}_X$ 的边界, 因此 $\bar{x}(u)$ 是 (P_u) 的局部最优解. $\square$.

注 5.18 在上述定理的假设下, 对所有 $u \in \mathcal{V}_U$, 问题 (P_u) 在 $\mathcal{V}_X$ 中有唯一的稳定点 $\bar{x}(u)$, $\bar{x}(u)$ 是 $f(\cdot, u)$ 在集合 $\{x \in \mathcal{V}_X : G(x,u) \in K\}$ 上的极小点. 由 (5.56) 还可以得到, 局部最优解 $\bar{x}(u) \in \mathcal{V}_X$ 关于 $u \in \mathcal{V}_U$ 是连续的且 $\bar{x}(u_0) = x_0$.

注 5.19 若所考虑的参数化满足约束映射与参数向量 u 独立, 则 (P_u) 的可行集是常值的, 根据命题 4.32 的上界结果, 由一致二阶增长条件可得, 存在 $\kappa > 0$, 满足

$$\|\bar{x}(u) - \bar{x}(u')\| \leqslant \kappa \|u - u'\|, \quad \forall u, u' \in \mathcal{V}_U, \tag{5.57}$$

即 $\bar{x}(u)$ 在 u_0 的邻域内是 Lipschitz 连续的. 尤其, 若关于倾斜参数化, 一致二阶增长条件在 x_0 处成立, 则对于在 $0 \in X^*$ 的某一邻域 $\mathcal{V}_{X^*}$ 的所有 δ_1, 存在 x_0 附近的唯一的稳定点 $\bar{x}(\delta_1)$, 它也是 (P_{δ_1}) 的局部最优解, $\bar{x}(\delta_1)$ 在 $\mathcal{V}_{X^*}$ 上是 Lipschitz 连续的.

下述定理讨论的是强正则性与一致二阶增长条件的联系.

定理 5.20 设 $(f(x,u), G(x,u))$ 是问题 (5.48) 的 C^2 光滑的参数化. 设

(i) (x_0, λ_0) 是广义方程 (5.47) 的强正则解.

(ii) x_0 是问题 (5.48) 的局部最优解.

(iii) 若 $(x,u) \to (x_0, u_0), \lambda(u) \in \Lambda(x,u)$, 则 $\lambda(u) \to \lambda_0$.

则存在 x_0 与 u_0 的邻域 $\mathcal{V}_X$ 与 $\mathcal{V}_U$, 满足

(a) 对任何 $u \in \mathcal{V}_U$, 参数化优化问题 (P_u) 有唯一临界点 $(\bar{x}(u), \bar{\lambda}(u)) \in \mathcal{V}_X \times Y^*$ 在 $\mathcal{V}_U$ 上是 Lipschitz 连续的.

(b) 对任何 $u \in \mathcal{V}_U$, 点 $\bar{x}(u)$ 是 (P_u) 的局部最优解.

(c) 一致二阶增长条件在 x_0 处成立.

相反地, 令 (x_0, λ_0) 是问题 (5.48) 的临界点, 设相对于标准参数化一致二阶增长条件在 x_0 处成立, 且 $DG(x_0)$ 是映上的, 则 (x_0, λ_0) 是广义方程 (5.47) 的强正则解.

证明 上述结论 (a) 由定理 5.13 可直接得到, 下证结论 (b). 注意到, (x_0, λ_0) 的强正则性可推出 Robinson 约束规范在 x_0 处成立. 下证在点 x_0 处, 由 (5.48) 给出的非扰动问题 (P_0) 的二阶增长条件成立. 用反证法. 令 $\hat{x} \neq x_0$ 是 (P_0) 的可行点, 满足 $f(\hat{x}) = f(x_0) + \alpha^2 \|\hat{x} - x_0\|^2$, 其中 $\alpha \geqslant 0$ 是某一非负数. 注意到, 根据假设 (ii), x_0 是 (P_0) 的局部最优解, 对充分接近于 x_0 的 $\hat{x}$, 这样的 α 是存在的, 则 $\hat{x}$ 是

(P_0) 的 ε 最优解, 其中 $\varepsilon := \alpha^2\|\hat{x}-x_0\|^2$. 由 Ekeland 变分原理 (定理 3.23) 可得, 存在 $\delta \in X^*$①与 (P_0) 的 ε 最优解 $\tilde{x}$, 满足 $\|\tilde{x}-\hat{x}\| \leqslant \alpha\|\hat{x}-x_0\|$, $\|\delta\| \leqslant \alpha\|\hat{x}-x_0\|$, 且 $\tilde{x}$ 是 $f(x)-\langle\delta,x\rangle$ 在约束 $G(x)\in K$ 下的极小化问题的稳定点. 若 $\hat{x}$ 接近于 x_0, α 充分小, 由上面的估计, $\tilde{x}$ 接近于 x_0, 由假设 (iii), 相应的 Lagrange 乘子 λ 也是这样. 由强正则性可得, 存在常数 $\gamma>0$, 满足对充分小的 $\alpha>0$ 与充分接近于 x_0 的 $\hat{x}$, $\|\tilde{x}-x_0\| \leqslant \gamma\|\delta\|$. 结果,

$$\|\hat{x}-x_0\| \leqslant \|\tilde{x}-x_0\| + \|\tilde{x}-\hat{x}\| \leqslant \alpha(\gamma+1)\|\hat{x}-x_0\|.$$

若 $\alpha<1$, 则有 $\hat{x}\neq x_0$, 由上述不等式有 $\alpha \geqslant (\gamma+1)^{-1}$, 因此 α 不会任意小. 这导致矛盾, 于是完成了问题 (P_0) 在 x_0 处的二阶增长条件的证明.

现在考虑一点 $u\in\mathcal{V}_U$. 设邻域 $\mathcal{V}_X$ 满足 (P_0) 的二阶增长条件在 $\mathcal{V}_X$ 上成立. 令 ε_k 是收敛到 0 的正数序列, x_k 是 $f(\cdot,u)$ 在集合 $\Psi(u):=\mathcal{V}_X\cap\Phi(u)$ 的 ε_k^2 极小点. 因为二阶增长条件与 Robinson 约束规范 在 x_0 处成立, 由命题 4.37(也可见命题 4.41), 对充分小的 ε_k, 如有必要, 可减小邻域 $\mathcal{V}_U$, 可设点 x_k 充分接近 x_0, 因而属于邻域 $\mathcal{V}_X$ 的内部. 由 Ekeland 变分原理, 存在 $x_k'\in X$, 满足 x_k' 是 $f(\cdot,u)$ 在 $\Psi(u)$ 上的 ε_k^2 极小点, $\|x_k-x_k'\|\leqslant\varepsilon_k$, 且 x_k' 是函数 $\phi_k(x):=f(x,u)+\varepsilon_k\|x-x_k'\|$ 在集合 $\Psi(u)$ 上的极小点. 因为对充分小的 ε_k, 可假设 x_k' 属于 $\mathcal{V}_X$ 的内部, 则对于满足 $\|\delta_k\|\leqslant\varepsilon_k$ 的某一 δ_k, x_k' 是 $f(\cdot,u)-\langle\delta_k,\cdot\rangle$ 在问题 (P_u) 的可行集 $\Phi(u)$ 上的极小化问题的稳定点. 由强正则性有 x_k' 是 Cauchy 列, 因而收敛到某一点 $\hat{x}$. 得到 $\hat{x}$ 是 $f(\cdot,u)$ 在 $\Psi(u)$ 上的极小点, 且是 V_X 的内部点, 因此是问题 (P_u) 的局部最优解. 结果, $\hat{x}$ 是 (P_u) 的稳定点, 因此与 $\bar{x}(u)$ 重合. 这完成了 (b) 的证明.

证明 (c). 因为 $\bar{x}(u)$ 是 $f(\cdot,u)$ 在集合 $\Psi(u)$ 上的极小点, 上述推证表明, (P_0) 在 x_0 处的二阶增长条件关于参数 u 的小的扰动是 "一致的", 所以, 可用完全相同的方式证明一致二阶增长条件成立.

我们证明相反的结论. 考虑标准的 (P_0) 的参数化, 相应的参数向量是 $\delta := (\delta_1,\delta_2)\in X^*\times Y$. 我们要证明对于 $\|\delta\|$ 充分小, 方程 (5.49) 在 (x_0,λ_0) 的邻域内有唯一解 $(h(\delta),\mu(\delta))$ 是 Lipschitz 连续的. 我们已经注意到, (5.49) 可解释为最优化问题 (5.50) 的最优性系统.

由一致二阶增长条件, 因为 $DG(x_0)$ 是映上的, 所以 Robinson 约束规范成立, 由定理 5.17 可推出, 存在 $0\in X^*\times Y$ 的邻域 $\mathcal{V}$ 和 x_0 的邻域 $\mathcal{V}_X$, 满足对所有的 $\delta\in\mathcal{V}$, (5.49) 有唯一解 $(h(\delta),\mu(\delta))$ 连续地依赖于 δ, 使得 $h(\delta)$ 是 (5.50) 的局部最优解. 剩下要证明映射 $(h(\cdot),\mu(\cdot))$ 是 Lipschitz 连续的. 因此, 考虑 $\mathcal{V}$ 中的两个元素 $\hat{\delta},\tilde{\delta}$, 与它们相联系的解记为 $(\hat{h},\hat{\mu})$ 与 $(\tilde{h},\tilde{\mu})$. 因为 $DG(x_0)$ 是映上的, 由开映射定理, 存在 $\bar{h}$ 满足 $DG(x_0)\bar{h}=\hat{\delta}_2-\tilde{\delta}_2$, $\|\bar{h}\|\leqslant M\|\hat{\delta}_2-\tilde{\delta}_2\|$, 其中 M 是不依赖 $\hat{\delta}$ 与 $\tilde{\delta}$ 的某个

① 原著为 $\delta\in X$.

常数. 作变量变换 $\eta := h - \bar{h}$ 后, 我们看到, 对 $\delta = \hat{\delta}$, 方程 (5.49) 等价于

$$D_{xx}^2 L(x_0, \lambda_0)\eta + DG(x_0)^*\mu = \hat{\delta}_1 - D_{xx}^2 L(x_0, \lambda_0)\bar{h},$$

$$G(x_0) + DG(x_0)\eta + \hat{\delta}_2 \in N_K^{-1}(\lambda_0 + \mu).$$

将上述问题解释为当 $\delta = \tilde{\delta}$ 时 (5.49) 的扰动, 其中扰动只进入到目标函数, 且是 $O(\|\hat{\delta} - \tilde{\delta}\|)$ 阶的. 由命题 4.32, 上述系统的解 η 满足 $\eta = \bar{h} + O(\|\hat{\delta} - \tilde{\delta}\|)$. 回到 $\bar{h}$ 的定义, 就得到 $\hat{h} = \tilde{h} + O(\|\hat{\delta} - \tilde{\delta}\|)$, 这就是期望的结果. □

推论 5.21 *设 x_0 是问题 (5.48) 的局部最优解. 设 $DG(x_0)$ 是映上的, λ_0 是相对应的 Lagrange 乘子. 则 (i) 一致二阶增长条件在 x_0 处成立的充分必要条件是点 (x_0, λ_0) 强正则的; (ii) 一致二阶增长条件在 x_0 处成立的充要条件是它相对于标准参数化亦成立.*

证明 假设 "$DG(x_0)$ 是映上的", 可推出定理 5.20 的假设 (iii). 因此, 结论 (i) 由定理 5.20 立即得到.

显然, 标准参数化是 C^2 光滑的参数化. 所以, 为了证明 (ii), 只需验证关于标准参数化的一致二阶增长条件意味着关于任何 C^2 光滑参数化的一致二阶增长条件. 由定理 5.20 可得, 若 $DG(x_0)$ 是映上的, 则关于标准参数化的一致二阶增长条件意味着 (x_0, λ_0) 处的强正则性, 进而意味着一致二阶增长条件. □

注 5.22 注意到, 强正则性可推出 Robinson 约束规范, 从而可推出 Lagrange 乘子的一致有界性. 若空间 Y 是有限维的, 则定理 5.20 的假设 (iii) 是假设 (i) 的结果.

注 5.23 有意思的是, 若强正则性在局部最优解处成立, 则由定理 5.20, 一致二阶增长条件成立, 因而由注 3.67 的推证, 作为临界锥子集的 X 的任何闭子空间, 均是可 Hilbert 化的. 尤其, 若无约束问题的一局部最优解是强正则的 (即存在且是局部 Lipschitz 连续的), 则空间是可 Hilbert 化的.

现在考虑集合 K 在 $y_0 := G(x_0)$ 处 C^2 简约为 Banach 空间Z 的闭集合 $\mathcal{C}$ 的情况, 即存在 y_0 的邻域 N 与二次连续可微映射 $\Xi : N \to Z$ 满足 $D\Xi(y_0) : Y \to Z$ 是映上的且 $K \cap N = \{y \in N : \Xi(y) \in \mathcal{C}\}$(见定义 3.135). 令 $\mathcal{G}(x) := \Xi(G(x))$ 且

$$\min_{x \in X} f(x) \quad \text{s.t} \quad \mathcal{G}(x) \in \mathcal{C} \tag{5.58}$$

是相应的简化问题. 因为 $D\Xi(y_0)$ 是映上的, 有它的伴随映射 $D\Xi(y_0)^* : Z^* \to Y^*$ 是一对一的且像等于 $[\ker D\Xi(y_0)]^\perp$, 因而是 Y^* 的闭子空间. 所以, 简化问题(5.58) 的临界点的强正则性与原问题 (5.48) 的临界点的强正则性是等价的. 在上述简化之下, 一致二阶增长条件也是被保持的.

点 x_0 相对于上述简化是非退化的当且仅当 $D\mathcal{G}(x_0)$ 是映上的 (见定义 4.70 及下述讨论). 由定理 5.20 可得, 若点 x_0 是非退化的且一致二阶增长条件在稳定点 x_0 处成立, 则 (x_0, λ_0) 是广义方程 (5.47) 的强正则解. 如下述定理所示, 若简约是点的, 则相反的结论亦成立.

定理 5.24　令 x_0 是问题 (5.48) 的局部最优解, λ_0 是相应的 Lagrange 乘子. 设空间 Y 是有限维的, 集合 K 在点 $G(x_0)$ 处 C^2 简约到一点的闭凸锥 $C \subset Z$. 则 (x_0, λ_0) 是广义方程 (5.47) 的强正则解的充要条件是 x_0 是非退化的且一致二阶增长条件在 x_0 处成立.

证明　由定理 5.20 可得, 若点 x_0 是非退化的且一致二阶增长条件成立, 则 (x_0, λ_0) 是强正则的. 相反地, 由定理 5.20, 若 (x_0, λ_0) 是强正则的, 则一致二阶增长条件成立. 所以只需要证明, 若 (x_0, λ_0) 是强正则的, 则 x_0 是非退化的.

先设严格互补条件 (定义 4.74) 在 x_0 处成立. 由命题 4.75, 有 Lagrange 乘子 λ_0 是唯一的当且仅当 x_0 是非退化的. 因为强正则性可推出 λ_0 的唯一性, 所以可得到 x_0 的非退化性质. 现在考虑一般情形. 因为 Y 是有限维的, 所以 $N_K(y_0)$ 具有非空的相对内部, 如它含有某一向量 $\mu \in Y^*$, 则 $\alpha := \mu - \lambda_0$ 满足 $\lambda_0 + t\alpha \in \mathrm{ri}(N_K(y_0))$ 对所有充分小的 $t > 0$ 成立. 将线性项 $\langle -t\alpha, DG(x_0)x\rangle$ 加到目标函数上, 即考虑标准形式的参数化问题 (P_δ), 其中 $\delta_1 := t[DG(x_0)]^*\alpha, \delta_2 := 0$, 则对于 $t > 0$ 充分小, 点 $(x_0, \lambda_0 + t\alpha)$ 是 (P_δ) 的稳定点, 且对问题 (P_δ), 严格互补条件在 $(x_0, \lambda_0 + t\alpha)$ 处成立. 由强正则性, $\lambda_0 + t\alpha$ 是唯一的, 从而 x_0 是非退化的. □

5.1.4　强正则性与二阶最优性条件

前面一节已证明, 局部最优解的强正则性与一致二阶增长条件是密切相联系的. 这一节给出一致二阶增长条件与二阶最优性条件间的联系. 首先给出一致二阶增长条件的必要条件.

令 x_0 是问题 (5.48) 的局部最优解, λ_0 是相应的 Lagrange 乘子, 令 $C(x_0)$ 是相应的临界锥, 可以表示为

$$C(x_0) = \{h \in X : DG(x_0)h \in T_K(y_0), \langle \lambda_0, DG(x_0)h\rangle = 0\}, \tag{5.59}$$

其中 $y_0 := G(x_0)$. 回顾雷达临界锥与强的广义多面性条件的定义 3.52. 由命题 3.54, 集合 $C_R(x_0)$ 是 $C(x_0)$ 的稠密子集合, 因而, 若 K 在 y_0 处是多面性的, $DG(x_0) : X \to Y$ 是映上的, 则强广义多面性条件是成立的.

定理 5.25　设 x_0 是最优化问题 (5.48) 的局部最优解, λ_0 是与 x_0 相联系的唯一的 Lagrange 乘子. 设 (相对于标准的参数化的) 一致二阶增长条件在 x_0 处成立. 则存在 $\alpha > 0$, 满足

$$D^2_{xx}L(x_0, \lambda_0)(h, h) \geqslant \alpha\|h\|^2, \quad \forall h \in \mathrm{Sp}[C_R(x_0)]. \tag{5.60}$$

若还有强的广义多面性条件成立, 则存在 $\alpha > 0$ 满足

$$D^2_{xx}L(x_0, \lambda_0)(h, h) \geqslant \alpha\|h\|^2, \quad \forall h \in \mathrm{Sp}[C(x_0)]. \tag{5.61}$$

证明 考虑点 $h_0 \in C_R(x_0)$, 标准的扰动 $\delta := (\delta_1, \delta_2)$, 其中 $\delta_1 := 0, \delta_2 := \tau DG(x_0)h_0$, $\tau > 0$ 满足

$$\bar{y} := G(x_0) + \tau DG(x_0)h_0 \in K. \tag{5.62}$$

则 (x_0, λ_0) 是临界点, 因而 x_0 是 $\tau > 0$ 充分小时的扰动问题 (P_δ) 的稳定点. 事实上, 由 (5.62) 可得 x_0 是 (P_δ) 的可行点, Lagrange 函数在 (x_0, λ_0) 处关于 x 的导数是零. 因为 h_0 是临界方向, 所以 $\langle \lambda_0, DG(x_0)h_0 \rangle = 0$. 因为 $\lambda_0 \in N_K(y_0)$, 所以对任何 $y \in K$, 有

$$\langle \lambda_0, y - \bar{y} \rangle = \langle \lambda_0, y - y_0 \rangle \leqslant 0,$$

因此 $\lambda_0 \in N_K(\bar{y})$.

令 $C_\delta(x_0)$ 是扰动问题 (P_δ) 在 x_0 处的临界锥. 考虑点 $h_1 \in C_R(x_0)$. 如有必要可减少 $\tau > 0$, 可假设 $\tilde{y} := G(x_0) + \tau DG(x_0)h_1 \in K$. 因为 $\langle \lambda_0, DG(x_0)h_0 \rangle = 0$, $\langle \lambda_0, DG(x_0)h_1 \rangle = 0$, $DG(x_0)(h_1 - h_0) = \tau(\tilde{y} - \bar{y})$, 得到

$$DG(x_0)(h_1 - h_0) \in T_K(\bar{y}) \cap \lambda_0^{\perp},$$

即 $h_1 - h_0 \in C_\delta(x_0)$. 因此, 由一致二阶增长条件结合定理 3.45 可推得

$$D^2_{xx}L(x_0, \lambda_0)(h_1 - h_0, h_1 - h_0) \geqslant \alpha\|h_1 - h_0\|^2,$$

其中 α 是不依赖于 h_1 与 h_0 的某正数. 证得 (5.60).

广义多面性条件意味着 $C_R(x_0)$ 是 $C(x_0)$ 的稠密子集, 于是得到 $\mathrm{Sp}[C_R(x_0)]$ 是 $\mathrm{Sp}[C(x_0)]$ 的稠密子集. 因为 $D^2_{xx}L(\cdot, \lambda_0)$ 是连续的, 所以 (5.61) 成立. □

若 $DG(x_0)$ 是映上的, 则由定理 5.20 与推论 5.21 可得强正则性等价于一致二阶增长条件, 现在给出二阶充分条件. 为继续讨论, 需要下述概念. 令 $K \subset Y$ 是闭凸锥, K^- 是其对偶空间 Y^* 中的极锥. 锥 K^- 在 Y^* 中引导一个序关系, 即 $y_1^* \preceq_* y_2^*$ 当且仅当 $y_2^* - y_1^* \in K^-$. 对两点 y_1^*, $y_2^* \in Y^*$, 相应的区间定义为

$$[y_1^*, y_2^*]_{K^-} := \{y^* \in Y^* : y_1^* \preceq_* y^* \preceq_* y_2^*\}.$$

这样的区间是弱 * 闭凸集合, 它是非空的当且仅当 $y_1^* \preceq_* y_2^*$. 用 $[y_1^*, y_2^*]_{K^-}^{\perp}$ 记满足 $\langle y^*, y \rangle = 0$, $\forall y^* \in [y_1^*, y_2^*]_{K^-}$ 的 $y \in Y$ 的集合.

引理 5.26 令 K 是闭凸锥, $y \in K, \lambda \in N_K(y)$. 则

(i) $[0,\lambda]_{K^-} \subset N_K(y)$.

(ii) 若还有 $z \in T_K(y) \cap \lambda^\perp$, 则 $[0,\lambda]_{K^-} \subset N_K(y) \cap z^\perp$.

(iii) 令 x_0 是最优化问题 (5.48) 的局部最优解, λ_0 是相应的 Lagrange 乘子. 则

$$\mathrm{cl}(\mathrm{Sp}(C(x_0))) \subset DG(x_0)^{-1}[0,\lambda_0]_{K^-}^\perp. \tag{5.63}$$

证明 (i) 由 (2.110), $N_K(x) = K^- \cap x^\perp$. 这一结果由例 2.62, 得到. 令 $\mu \in [0,\lambda]_{K^-}$. 因为 $y \in K$, 有 $\langle \lambda - \mu, y\rangle \leqslant 0$. 又由 $\langle \lambda, y\rangle = 0$, 得到 $\langle \mu, y\rangle \geqslant 0$. 然而 $\mu \in K^-$, 因此有 $\langle \mu, y\rangle = 0$, 证得 (i).

(ii) 因为 $\mu \in [0,\lambda]_{K^-}$, 由结论 (i), μ 属于 $N_K(y)$, 有 $\langle \mu, z\rangle \leqslant 0$. 另一方面, $T_K(y) = \mathrm{cl}\{K + [\![y]\!]\}$, 因而有 $z = \lim_{n\to\infty}(z_n + t_n y)$, 其中 $z_n \in K$ 是某一序列, $t_n \in \mathbb{R}$. 所以,

$$-\langle \mu, z\rangle = \langle \lambda - \mu, z\rangle = \lim_{n\to\infty}\langle \lambda - \mu, z_n\rangle \leqslant 0,$$

因此 $\langle \mu, z\rangle = 0$, 证得 (ii).

(iii) 令 $\mu \in [0,\lambda_0]_{K^-}$. 由 (ii), 对每一临界方向 h, $\langle \mu, DG(x_0)h\rangle = 0$, 即 $C(x_0) \subset DG(x_0)^{-1}[0,\lambda_0]_{K^-}^\perp$, 由此可得 (iii). □

现在考虑 $(Y^*, \preceq_*)$ 具有格结构(定义 3.56) 的情形.

定理 5.27 令 x_0 是最优化问题 (5.48) 的局部最优解, λ_0 是与 x_0 相联系的 Lagrange 乘子, 即 (x_0, λ_0) 是 (5.48) 的临界点. 设

(i) 空间 X 是自反的.

(ii) $K \subset Y$ 是闭凸锥, 且 $(Y^*, \preceq_*)$ 具有格结构.

(iii) $DG(x_0): X \to Y$ 是映上的.

(iv) $Q(h) := D^2_{xx}L(x_0,\lambda_0)(h,h)$ 是 Legendre 形式.

(v) 下述二阶条件成立:

$$D^2_{xx}L(x_0,\lambda_0)(h,h) > 0, \quad \forall h \neq 0, \quad DG(x_0)h \in [0,\lambda_0]_{K^-}^\perp. \tag{5.64}$$

则一致二阶增长条件在 x_0 处成立.

证明 用反证法. 证明分为两步. 步 (a) 不用由 K^- 引导的格结构, 为叙述定理的某些变形, 可参看下面的注 5.31 和注 5.32.

(a) 设一致二阶增长条件不成立, 则存在 C^2 光滑的参数化函数 $(f(x,u), G(x,u))$ 与序列 $u_n \to u_0$, $x_n \to x_0$, $h_n \to 0$ 满足 x_n 与问题 (P_{u_n}) 的 Lagrange 乘子 λ_n 相联系, $G(x_n + h_n, u_n) \in K$ 且

$$f(x_n + h_n, u_n) \leqslant f(x_n, u_n) + o(\|h_n\|^2). \tag{5.65}$$

由 $DG(x_0)$ 是映上的, 有 $\lambda_n \to \lambda_0$. 进一步, 由于 X 是自反的, 如有必要, 可取一子序列, 设 $\bar{h}_n := h_n/\|h_n\|$ 弱收敛到某一向量 $\bar{h}$. 由 $\lambda_n \in N_K(G(x_n, u_n))$ 且 $G(x_n + h_n, u_n) \in K$, 有

$$\langle \lambda_n, G(x_n + h_n, u_n) - G(x_n, u_n) \rangle \leqslant 0, \tag{5.66}$$

结合 (5.65) 可推出

$$\begin{aligned} L(x_n + h_n, \lambda_n, u_n) - L(x_n, \lambda_n, u_n) &\leqslant f(x_n + h_n, u_n) - f(x_n, u_n) \\ &\leqslant o(\|h_n\|^2). \end{aligned}$$

由 Lagrange 函数的 Hesse 阵是弱下半连续的, 有

$$\begin{aligned} Q(\bar{h}) = D_{xx}^2 L(x_0, \lambda_0)(\bar{h}, \bar{h}) &\leqslant \liminf_{n\to\infty} D_{xx}^2 L(x_0, \lambda_0)(\bar{h}_n, \bar{h}_n) \\ &= \liminf_{n\to\infty} D_{xx}^2 L(x_n, \lambda_n)(\bar{h}_n, \bar{h}_n) \\ &= 2 \inf\lim_{n\to\infty} \|h_n\|^{-2}(L(x_n + h_n, \lambda_n, u_n) - L(x_n, \lambda_n, u_n)), \end{aligned}$$

从而有

$$D_{xx}^2 L(x_0, \lambda_0)(\bar{h}, \bar{h}) \leqslant 0, \tag{5.67}$$

$\bar{h}$ 不会是 0, 因为它是单位范数向量的弱极限且 $Q(\cdot)$ 是 Legendre 形式.

由 (5.65), 有 $D_x f(x_n, u_n)h_n \leqslant o(\|h_n\|)$, 由最优性系统可得

$$\langle \lambda_n, D_x G(x_n, u_n)h_n \rangle \geqslant o(\|h_n\|). \tag{5.68}$$

另一方面, 由 (5.66) 可得

$$\begin{aligned} \langle \lambda_n, D_x G(x_n, u_n)h_n \rangle &= \langle \lambda_n, G(x_n + h_n, u_n) - G(x_n, u_n) \rangle + o(\|h_n\|) \\ &\leqslant o(\|h_n\|). \end{aligned}$$

得到 $\langle \lambda_n, D_x G(x_n, u_n)h_n \rangle = o(\|h_n\|)$, 因此

$$\langle \lambda_0, D_x G(x_0, u_0)\bar{h} \rangle = 0. \tag{5.69}$$

这一步的结论是, 在除了格结构外的定理的假设下, 存在非零的方向 $\bar{h}$, 满足 (5.69) 与 (5.67).

(b) $\lambda_n \in K^-$ 即 $0 \preceq_* \lambda_n$, 令 $\mu \in [0, \lambda_0]$, 则 $\mu_n := \mu \wedge \lambda_n$ 属于 $[0, \lambda_n]$, 且由 “下确界” 算子 $\wedge$ 的连续性, μ_n 收敛到 $\mu \wedge \lambda_0 = \mu$. 由引理 5.26(i) 可推出 $\mu_n \in N_K(G(x_n, u_n))$, 因而 $\langle \lambda_n, G(x_n, u_n) \rangle = \langle \mu_n, G(x_n, u_n) \rangle = 0$, 与 $G(x_n + h_n, u_n) \in K$ 相结合可得

$$\begin{aligned} \langle \mu_n, D_x G(x_n, u_n)h_n \rangle &= \langle \mu_n, G(x_n + h_n, u_n) \rangle + o(\|h_n\|) \\ &\leqslant o(\|h_n\|). \end{aligned} \tag{5.70}$$

因为 $\mu_n \in [0,\lambda_n]$, $G(x_n+h_n,u_n)\in K$, 根据 (5.68) 得到

$$\begin{aligned}\langle \mu_n, D_xG(x_n,u_n)h_n\rangle &= \langle \mu_n, G(x_n+h_n,u_n)-G(x_n,u_n)\rangle + o(\|h_n\|)\\ &\geqslant \langle \lambda_n, G(x_n+h_n,u_n)\rangle + o(\|h_n\|)\geqslant o(\|h_n\|),\end{aligned}$$

把上式与 (5.70) 相结合可推出 $\langle \mu_n, D_xG(x_n,u_n)h_n\rangle = o(\|h_n\|)$, 取极限可得 $\langle \mu, D_xG(x_0,u_0)\bar{h}\rangle = 0$. 因为此式对任何 $\mu\in[0,\lambda_0]$ 均成立, 所以得到 $D_xG(x_0,u_0)\bar{h}\in [0,\lambda_0]^{\perp}$. 由 (5.64) 与 (5.69) 可推出 $\bar{h}=0$, 导致矛盾. □

注 5.28　因为我们需要的是与 x_n 相联系的 Lagrange 乘子 λ_n 存在且收敛到 λ_0, 所以可以减弱假设 (iii)(指 $DG(x_0)$ 是映上的). 例如, 若 Y 是有限维的, 非扰动问题具有唯一的 Lagrange 乘子 (因此 Robinson 约束规范在 x_0 处成立), 则结论仍然成立.

推论 5.29　设定理 5.27 中的假设成立. 则问题 (5.48) 的临界点(x_0,λ_0) 是强正则的.

证明　由定理 5.20, 又因 $DG(x_0)$ 是映上的, 所以由一致二阶增长条件可得 (x_0,λ_0) 的强正则性. □

现在给出强正则性的下述刻画.

定理 5.30　设

(i) X 是自反的空间.

(ii) K 是闭凸锥且 $(Y^*,\preceq_*)$ 具有格结构.

(iii) x_0 是最优化问题 (5.48) 的一个局部最优解.

(iv) $DG(x_0)$ 是映上的, λ_0 记相应的 (唯一的)Lagrange 乘子.

(v) 强广义多面性条件成立.

(vi) $Q(h):=D^2_{xx}L(x_0,\lambda_0)(h,h)$ 是 Legendre 形式.

(vii) 下述等式成立

$$\mathrm{cl}(\mathrm{Sp}\ (C(x_0))) = DG(x_0)^{-1}[0,\lambda_0]^{\perp}_{K^-}. \tag{5.71}$$

则临界点(x_0,λ_0) 是强正则的充分必要条件是二阶条件 (5.64) 成立.

证明　由推论 5.29 结合定理 5.20 的最后结论, 有 (5.64) 是强正则性的充分条件. 相反地, 设强正则性成立. 由定理 5.20, 有一致二阶增长条件是满足的. 因为强广义多面性的条件成立, 所以由定理 5.25 得到, 存在 $\alpha>0$, (5.61) 成立. 由假设 (5.71) 可推出条件 (5.64). □

注 5.31　可以给出定理 5.27 的另一个可能的推广. 约束具有下述乘积的形式: $Y=Y_1\times Y_2$, $G(x)=(G_1(x),G_2(x))\in Y_1\times Y_2$, $K=\{0\}\times K_2$, $K_2\subset Y_2$ 是闭凸

锥, 满足当赋予由 K_2^- 引导的序关系时 Y_2^* 是 Banach 格. 把 (5.64) 变成

$$\begin{cases} D_{xx}^2 L(x_0,\lambda_0)(h,h) > 0, \quad \forall h \neq 0, \\ DG_1(x_0)h = 0, \quad DG_2(x_0)h \in [0,\lambda_{02}]_{K_2^-}^{\perp}, \end{cases} \tag{5.72}$$

就可得一致二阶增长条件的一个充分条件. 这里 λ_{02} 记 Lagrange 乘子的第二分量.

注 5.32 对于约束具有注 5.31 所述的乘积形式但没有格结构的情形, 可以给出定理 5.27 的另一有用的变形. 此种情形, 可得 (5.64) 变为

$$\begin{cases} D_{xx}^2 L(x_0,\lambda_0)(h,h) > 0, \quad \forall h \neq 0, \\ DG_1(x_0)h = 0, \quad \langle \lambda_0, DG(x_0)h \rangle = 0. \end{cases} \tag{5.73}$$

为得到这一结果, 观察到定理 5.27 证明中的步 (a) 没有用到格结构, 而这一步中构造的方向 $\bar{h}$ 与 (5.73) 矛盾即可.

5.1.5 强稳定性

这一节研究稳定点与局部最优解的稳定性(连续性), 不要求将之与相应的 Lagrange 乘子一起处理. 事实上, 在某些情况下, 即使相对应的 Lagrange 乘子不是连续的, 局部最优解也是稳定的, 甚至是 Lipschitz 连续的 (例如, 在半无限规划的框架下, Lagrange 乘子由相应指标集上的测度给出就是这种情况, 见注 5.120). 还可以证明, 局部最优解的稳定性与一致二阶增长条件是密切相关的.

考虑由 (5.48) 给出的 (非扰动) 最优化问题 (P_0).

定义 5.33 称问题 (P_0) 的稳定点 x_0 关于 (P_0) 的 C^2 光滑参数化 $(f(x,u),G(x,u))$ 是强稳定的 (强 Lipschitz 稳定的), 若存在 x_0 与 u_0 的邻域 $\mathcal{V}_X$ 与 $\mathcal{V}_U$, 满足对任何的 $u\in\mathcal{V}_U$, 相应的问题 (P_u) 有唯一的稳定点 $\bar{x}(u)\in\mathcal{V}_X$, 且映射 $u\to\bar{x}(u)$ 在 $\mathcal{V}_U$ 上是连续的 (Lipschitz 连续的). 若对 (P_0) 的任何 C^2 光滑参数化均是成立的, 则称 x_0 是强稳定的 (强 Lipschitz 稳定的).

注意到, 在规范与二阶增长条件成立的前提下, 若空间 X 是有限维的, x_0 是 (P_0) 的局部最优解, 则由紧致性推证可得, 稳定点 $\bar{x}(u)$ 的连续性可由唯一性假设自动得到.

下述定理表明, 强稳定性与一致二阶增长条件是密切联系的.

定理 5.34 设 x_0 是 (P_0) 的局部最优解, Robinson 约束规范在 x_0 处成立 (因此 x_0 是 (P_0) 的稳定点), 则 x_0 处的一致二阶增长条件成立的充分必要条件是 x_0 是强稳定的.

证明 由定理 5.17, 若一致二阶增长条件在 x_0 处成立, 则对与 u_0 充分接近的所有的 u, 问题 (P_u) 在 x_0 的邻域内具有稳定点 $\bar{x}(u)$, 且 $\bar{x}(u)$ 是 Hölder 连续的 (见 (5.56)), 因而是唯一的.

相反地, 设 x_0 是强稳定的. 先证明二阶增长条件在 x_0 处成立. 因为 x_0 是 (P_0) 的局部最优解, 存在 x_0 的邻域 $\mathcal{V}_X$, 满足 x_0 是 $f(x)$ 在 $\mathcal{V}_X \cap G^{-1}(K)$ 上的极小点. 考虑点 $\hat{x} \in \mathcal{V}_X \cap G^{-1}(K), \hat{x} \neq x_0$, 则存在 $\alpha \geqslant 0$, 满足 $f(\hat{x}) = f(x_0) + \alpha^2\|\hat{x} - x_0\|^2$, 结果 $\hat{x}$ 是 (P_0) 限定在 $\mathcal{V}_X$ 上的 ε 最优解, 其中 $\varepsilon := \alpha^2\|\hat{x} - x_0\|^2$. 由 Ekeland 变分原理 (定理 3.22), 存在 $\delta \in X^*$①及 (P_0) 的 ε 最优解 $\tilde{x}$, 满足 $\|\tilde{x} - \hat{x}\| \leqslant \alpha\|\hat{x} - x_0\|$, $\|\delta\| \leqslant \alpha\|\hat{x} - x_0\|$, 且 $\tilde{x}$ 是 $f(x) - \langle\delta, x\rangle$ 受约束于 $G(x) \in K$ 的极小化问题的稳定点. 于是, 若 $\alpha < 1$, 则 $\|\hat{x} - x_0\| \leqslant (1-\alpha)^{-1}\|\tilde{x} - x_0\|$, 因而有 $\|\delta\| \leqslant \alpha(1-\alpha)^{-1}\|\tilde{x} - x_0\|$.

现在证明存在线性连续的自伴随算子 $A : X \to X^*$, 满足 $\delta = A(\tilde{x} - x_0)$ 且 $\|A\| \leqslant 3\alpha(1-\alpha)^{-1}$. 事实上, 考虑 $h := \tilde{x} - x_0$. 由 Hahn-Banach 定理, 存在 $h^* \in X^*$ 满足 $\|h^*\| = 1$, $\langle h^*, h\rangle = \|h\|$. 定义

$$Ax := \|h\|^{-1}[\langle h^*, x\rangle\delta + \langle\delta, x\rangle h^*] - \|h\|^{-2}\langle\delta, h\rangle\langle h^*, x\rangle h^*. \tag{5.74}$$

显然, A 是自伴随的, $Ah = \delta$, 且

$$\|A\| \leqslant 2\|h\|^{-1}\|h^*\|\|\delta\| + \|h\|^{-1}\|\delta\|\|h^*\|^2 = 3\|h\|^{-1}\|\delta\| \leqslant 3\alpha(1-\alpha)^{-1}.$$

于是得 x_0 与 $\tilde{x}$ 是函数 $f(x) - \frac{1}{2}\langle x - x_0, A(x - x_0)\rangle$ 在 $G^{-1}(K)$ 上的极小化问题的稳定点. 考虑 (P_0) 的参数化, 其中 $U := \mathcal{L}(X, X^*), f(x,u) := f(x) + \langle x - x_0, u(x - x_0)\rangle$ 且 $G(x,u) := G(x)$. 因而, 若 α 是任意小的, 则对任意接近于 $0 \in U$ 的某一 $u \in U$, 相应的问题 (P_u) 在 x_0 的邻域内有两个不同的稳定点. 然而, 这与 x_0 的强稳定性矛盾. 于是证得 x_0 处的二阶增长条件成立.

再者, 由 (P_0) 的 C^2 光滑参数化 (P_u) 的稳定点 $\bar{x}(u)$ 的连续性, 用与定理 5.20 证明相同的推证, 可以证明 $\bar{x}(u)$ 是 $f(\cdot, u)$ 在 $\mathcal{V}_X \cap \Phi(u)$ 上的极小点, 其中 $\mathcal{V}_X$ 是 x_0 的邻域, $\Phi(u)$ 是 (P_u) 的可行集, 则可以用与上述在 x_0 处的二阶增长条件相同的推证完成一致二阶增长条件的证明. □

上述定理与定理 5.24 可推出强正则性与强稳定性这两个概念间的关系.

定理 5.35　*设 x_0 是 (P_0) 的局部最优解, λ_0 是相应的 Larange 乘子. 设*

(i) 空间 Y 是有限维的.

(ii) 集合 K 在点 $G(x_0)$ 处 C^2 简约为点的闭凸锥.

(iii) 点 x_0 关于这一简约是非退化的.

则 (x_0, λ_0) 是广义方程 (5.47) 的强正则解的充分必要条件是 x_0 为 (P_0) 的强稳定点.

考虑倾斜参数化, 即只有目标函数被参数化为 $f(x, \delta_1) = f(x) - \langle\delta_1, x\rangle, \delta_1 \in X^*$ 的情况. 已经证得, 若一致二阶增长条件在 x_0 关于倾斜参数化成立, 则 x_0 关于倾

① 原著中为 $\delta \in X$.

斜参数化是强 Lipschitz 稳定的 (见注 5.19). 在 Robinson 约束规范条件成立的前提下, 相反的结论也是成立的.

定理 5.36 设 x_0 是 (P_0) 的局部最优解, 则 x_0 相对于倾斜参数化是强 Lipschitz 稳定的当且仅当相对于倾斜参数化一致二阶增长条件在 x_0 处成立.

证明 一个推出关系在注 5.19 中已经证得. 下证相反的推出关系也是成立的. 设 x_0 相对于倾斜参数化是强 Lipschitz 稳定的.

因为 x_0 是 (P_0) 的局部最优解, 由定理 5.34 的证明可得, 存在 x_0 与 $0 \in X^*$ 的邻域 $\mathcal{V}_X$ 与 $\mathcal{V}_U$, 满足对任何 $\delta_1 \in \mathcal{V}_U$, 相应的稳定点 $\bar{x}(\delta_1)$ 是 $f(\cdot,\delta_1) := f(\cdot) - \langle \delta_1, \cdot \rangle$ 在 $\mathcal{V}_X$ 上的极小点, 进一步, $\bar{x}(\delta_1)$ 是 $\mathcal{V}_X$ 的内部点.

对给定的 $\delta_1 \in \mathcal{V}_U$, 考虑与稳定点 $\bar{x} = \bar{x}(\delta_1)$ 不同的一点 $\hat{x} \in \mathcal{V}_X \cap G^{-1}(K)$. 因 $\bar{x}$ 是 $f(\cdot,\delta_1)$ 在 $\mathcal{V}_X$ 上的极小点, 存在 $\alpha \geqslant 0$, $f(\hat{x},\delta_1) = f(\bar{x},\delta_1) + \alpha^2\|\hat{x} - \bar{x}\|^2$, 因此 $\hat{x}$ 是 $f(\cdot,\delta_1)$ 在 $\mathcal{V}_X$ 上的 ε 极小点, 其中 $\varepsilon := \alpha^2\|\hat{x} - \bar{x}\|^2$. 由 Ekeland 变分原理, 存在 $\delta' \in X^*$ (译者注: 原著中为 $\delta' \in X$), 以及 $f(\cdot,\delta_1)$ 在 $\mathcal{V}_X$ 上的 ε 极小点 $\tilde{x}$, 满足 $\|\tilde{x} - \hat{x}\| \leqslant \alpha\|\hat{x} - \bar{x}\|$, $\|\delta'\| \leqslant \alpha\|\hat{x} - \bar{x}\|$, 且 $\tilde{x}$ 是函数 $f(\cdot,\delta_1) + \langle \delta', \cdot \rangle$ 的稳定点. 由强 Lipschitz 稳定性可得, 存在常数 $\gamma > 0$, 满足对充分小的 $\alpha > 0$ 与充分接近于 $\bar{x}$ 的 $\hat{x}$, 有 $\|\tilde{x} - \bar{x}\| \leqslant \gamma\|\delta'\|$. 结果,

$$\|\hat{x} - \bar{x}\| \leqslant \|\tilde{x} - \bar{x}\| + \|\tilde{x} - \hat{x}\| \leqslant \alpha(\gamma + 1)\|\hat{x} - \bar{x}\|.$$

若 $\alpha < 1$, 则 $\|\hat{x} - \bar{x}\| \neq 0$. 于是有 $\alpha \geqslant (\gamma + 1)^{-1}$, α 不会任意小. 这就完成了相应的二阶增长条件的证明. □

5.1.6 一些例子及应用

这一节给出上述抽象结果的某些应用.

1. 非线性规划

考虑非线性规划问题

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x), \\ \text{s. t.} \quad & g_i(x) = 0, \quad i = 1, \cdots, q, \\ & g_i(x) \leqslant 0, \quad i = q+1, \cdots, p, \end{aligned} \tag{5.75}$$

其中 $f, g_1, \cdots, g_p$ 是二次连续可微的.

令 x_0 是局部最优解, $\overline{\lambda} \in \Lambda(x_0)$ 是非线性规划问题 (5.75) 相应的 Larange 乘子. 考虑下述指标集: 在 x_0 处起作用的不等式约束指标集 $I(x_0)$ 与

$$\begin{aligned} J(x_0) := \{1, \cdots, q\} \cup I(x_0), \quad & I_+(x_0, \overline{\lambda}) := \{i \in I(x_0) : \overline{\lambda}_i > o\}, \\ I_0(x_0, \overline{\lambda}) := \{i \in I(x_0) : \overline{\lambda}_i = 0\} & = I(x_0) \backslash I_+(x_0, \overline{\lambda}). \end{aligned} \tag{5.76}$$

对任何 $\overline{\lambda} \in \Lambda(x_0)$, 在 x_0 处的临界锥可以表示为

$$C(x_0) = \Big\{h : Dg_i(x_0)h = 0,\ \ i \in \{1, \cdots, q\},\ \ Dg_i(x_0)h \leqslant 0, i \in I(x_0),\\ \sum_{i \in J(x_0)} \overline{\lambda}_i Dg_i(x_0)h = 0\Big\},$$

或等价地

$$C(x_0) = \{h : Dg_i(x_0)h = 0,\ \ i \in \{1, \cdots, q\} \cup I_+(x_0, \overline{\lambda}),\\ Dg_i(x_0)h \leqslant 0, i \in I_0(x_0, \overline{\lambda})\}. \tag{5.77}$$

Lagrange 乘子 λ_0 是唯一的, 即 $\Lambda(x_0) = \{\lambda_0\}$ 当且仅当 (4.124) 成立. 在这些条件下, 有

$$\mathrm{Sp}\ [C(x_0)] = \{h : Dg_i(x_0)h = 0, i \in \{1, \cdots, q\} \cup I_+(x_0, \lambda_0)\}. \tag{5.78}$$

事实上, $\mathrm{Sp}[C(x_0)]$ 包含 (5.78) 的右端是 (5.77) 的一个显然的结论. 相反地, (5.78) 右端的每一 h 都可以表示为 $h = \alpha\bar{h} - (\alpha\bar{h} - h)$, 其中 $\alpha \geqslant 0$; 对充分大的 α, $(\alpha\bar{h} - h)$ 是临界方向, 因此 $h \in \mathrm{Sp}\ [C(x_0)]$. 于是证得, 若 $\Lambda(x_0) = \{\lambda_0\}$, 则 (5.78) 成立.

若在 x_0 处 Mangasarian-Fromovitz 约束规范成立, 则 Lagrange 乘子的集合 $\Lambda(x_0)$ 是非空的凸有界的 (因此是紧致的) 多胞体. 结果, $\Lambda(x_0)$ 与它的极点的凸包是重合的. 注意到, $\overline{\lambda} \in \Lambda(x_0)$ 是 $\Lambda(x_0)$ 的极点的充分必要条件是梯度向量 $\nabla g_i(x_0), i \in \{i, \cdots, q\} \cup I_+(x_0, \overline{\lambda})$ 是线性无关的.

现在考虑如下的二阶条件:

对任何 $\overline{\lambda} \in \Lambda(x_0)$ 及满足 $Dg_i(x_0)h = 0,\ i \in \{1, \cdots, q\} \cup I_+(x_0, \overline{\lambda})$ 的 $h \neq 0$,

$$D^2_{xx}L(x_0, \overline{\lambda})(h, h) > 0. \tag{5.79}$$

注意到, 若 Mangasarian-Fromovitz 约束规范在 x_0 处成立, 则任何 Lagrange 乘子 $\overline{\lambda} \in \Lambda(x_0)$ 均可表示为 $\Lambda(x_0)$ 的极点的凸组合, 则 $I_+(x_0, \overline{\lambda})$ 为这些极点相对应的指标集的并, 因此只需验证上述条件 (5.79) 对极点 $\overline{\lambda} \in \Lambda(x_0)$ 成立即可.

命题 5.37　*令 x_0 是非线性规划问题 (5.75) 的局部最优解, 设 Mangasarian-Fromovitz 约束规范在 x_0 处成立, 则下述结论是等价的:*

(i) 点 x_0 是强稳定的.

(ii) 一致二阶增长条件在 x_0 处成立.

(iii) 二阶条件 (5.79) 成立.

证明　(i) 与 (ii) 的等价性由定理 5.34 证得. 下证 (ii) 与 (iii) 的等价性.

设一致二阶增长条件在 x_0 处成立. 令 $\overline{\lambda} \in \Lambda(x_0)$ 是 $\Lambda(x_0)$ 的极点. 在不等式约束 $g_i(x) \leqslant 0$, $i \in I_0(x_0, \overline{\lambda})$ 的右端作小的扰动, 可以使这些约束不起作用. 在这样的一个扰动下, $(x_0, \overline{\lambda})$ 仍然是扰动问题的临界点, 因而对于扰动问题在 x_0 处的一致二阶增长条件亦是成立的. 因为 $\nabla g_i(x_0), i \in \{i, \cdots, q\} \cup I_+(x_0, \overline{\lambda})$ 是线性无关的, 由定理 5.25 可得, (相对于标准参数化) 在 x_0 处的一致二阶增长条件可推出条件 (5.79).

相反地, 设条件 (5.79) 成立. 用反证法. 假设一致二阶增长条件不成立, 则存在 (5.75) 的 C^2 光滑的参数化和序列 $u_n \to u_0$, $x_n \to x_0$, $h_n \to 0$, $h_n \neq 0$, 满足 x_n 与 $x_n + h_n$ 是 (P_{u_n}) 的可行点, x_n 是稳定点, 因此, 存在与 x_n 相联系的 Larange 乘子 λ_n, 且

$$f(x_n + h_n, u_n) \leqslant f(x_n, u_n) + o(\|h_n\|^2). \tag{5.80}$$

得到 $g_i(x_n, u_n) = 0$, $i \in \{i, \cdots, q\} \cup I_+(x_n, \lambda_n)$. 因为 $x_n \to x_0, \mathrm{dist}(\lambda_n, \Lambda(x_0)) \to 0$, $\Lambda(x_0)$ 具有有限多个极点, 如有必要, 可取一子列, 设对某一极点$\overline{\lambda} \in \Lambda(x_0)$, 有

$$g_i(x_n, u_n) = 0, \quad i \in \{i, \cdots, q\} \cup I_+(x_0, \overline{\lambda}),$$

且 $h_n/\|h_n\|$ 收敛到向量 $\bar{h}$. 因为 $x_n + h_n$ 是 (P_{u_n}) 的可行点, 因此满足相应的约束, 有

$$Dg_i(x_0)\bar{h} = 0, \quad i = 1, \cdots, q, \quad Dg_i(x_0)\bar{h} \leqslant 0, \quad i \in I_+(x_0, \overline{\lambda}). \tag{5.81}$$

由 (5.80) 还可以得到 $Df(x_0)\bar{h} \leqslant 0$. 结果,

$$-Df(x_0)\bar{h} = \sum_{i \in \{1, \cdots, q\} \cup I_+(x_0, \overline{\lambda})} \overline{\lambda}_i Dg_i(x_0)\bar{h} \geqslant 0,$$

结合 (5.81) 可推出

$$Dg_i(x_0)\bar{h} = 0, \quad i \in \{1, \cdots, q\} \cup I_+(x_0, \overline{\lambda}). \tag{5.82}$$

得到

$$f(x_n + h_n, u_n) - f(x_n, u_n) \geqslant L(x_n + h_n, u_n, \lambda_n) - L(x_n, u_n, \lambda_n),$$

因此, 由于 (5.82), 由条件 (5.79) 可得

$$f(x_n + h_n, u_n) - f(x_n, u_n) \geqslant \alpha \|h_n\|^2,$$

其中 $\alpha > 0$. 然而, 最后一不等式与 (5.80) 矛盾. 证毕. □

称在可行点 x_0 处的线性无关约束规范成立, 若梯度向量 $\nabla g_i(x_0), i \in J(x_0)$ 是线性无关的. 注意到, 线性无关的约束规范可推出 $\Lambda(x_0) = \{\lambda_0\}$ 是单点集, 从而条件 (5.79) 具有下述形式

对任何满足 $Dg_i(x_0)h = 0$, $i \in \{1, \cdots, q\} \cup I_+(x_0, \lambda_0)$ 的 $h \neq 0$,均有

$$D^2_{xx}L(x_0, \lambda_0)(h, h) > 0. \tag{5.83}$$

命题 5.38　设 x_0 是局部最优解, λ_0 是非线性规划问题 (5.75) 的相应的 Lagrange 乘子, 则临界点(x_0, λ_0) 是强正则的充分必要条件是线性无关的约束规范成立且二阶条件 (5.38) 成立.

证明　观察到, 非线性规划问题 (5.75) 是 (5.48) 的特例, 其中 $X := \mathbb{R}^n, Y := \mathbb{R}^p, K := \{0\} \times \mathbb{R}_-^{p-q}$ 且集合 K 是 C^2 简化到一点锥的. 所以, 由定理 5.24 得, 强正则性条件成立的充分必要条件是点 x_0 是非退化的且一致二阶增长条件成立. 当前的情形, 非退化性等价于线性无关的约束规范. 应用命题 5.37 即可完成证明.　□

注 5.39　由定理 5.34, 在 Mangasarian-Fromovitz 约束规范下, 局部最优解 x_0 是强稳定的当且仅当一致二阶增长条件成立. 由命题 5.37, 一致二阶增长条件等价于条件 (5.79). 由条件 (5.79) 可推得, 对 (5.75) 的任何 C^2 光滑参数化, 相应的局部最优解 $\bar{x}(u)$ 在 u_0 处是 Lipschitz 稳定的 (见定理 4.51), 由定理 4.95(见注 4.98 与 5.2.3 节中的直接分析) 得 $\bar{x}(u)$ 在 u_0 处是方向可微的, 其方向导数 $\bar{x}'(u_0, d)$ 由下述问题的最优解 $\bar{h} = \bar{h}(d)$ 给出

$$\min_{h \in \mathcal{S}(\mathrm{PL}d)} \Big\{ \max_{\lambda \in \mathcal{S}(\mathrm{DL}d)} D^2 L(x_0, \lambda, u_0)((h, d), (h, d)) \Big\}. \tag{5.84}$$

2. 平方可积函数空间

令 $(\Omega, \mathcal{F}, \mu)$ 是测度空间, $Y := L_2(\Omega, \mathcal{F}, \mu) = L_2(\Omega)$ 是相应的平方可积函数构成的 Hilbert 空间 (见例 2.39). 考虑到 Y^* 与 Y 相同, 且对 y 与 $\lambda \in Y$, y 与 λ 的数积由 $\int_\Omega y(\omega)\lambda(\omega)\mathrm{d}\mu(\omega)$ 给出. 令 $K := [L_2(\Omega)]_-$ 为几乎处处非正值函数构成的锥. K 的极锥是几乎处处 (a.e.) 非负值函数的锥 $K^- = [L_2(\Omega)]_+$. 两个函数 y_1, $y_2 \in L_2(\Omega)$ 间由 $[L_2(\Omega)]_+$ 引导的序关系 $y_1 \preceq_* y_2$ 意味着 $y_1(\omega) \leqslant y_2(\omega)$ 对几乎处处的 $\omega \in \Omega$ 成立. 这一序关系具有格结构, 相应的上确界 $y_1 \vee y_2$ 与下确界 $y_1 \wedge y_2$ 分别由 y_1 与 y_2 的几乎处处的逐点极大与极小给出.

引理 5.40　设 x_0 是最优化问题 (5.48) 的局部最优解, $Y := L_2(\Omega)$, $K := [L_2(\Omega)]_-$, 满足 $DG(x_0)$ 是映上的, 令 λ_0 是与 x_0 相联系的 Lagrange 乘子. 则 (5.71) 成立, 且

$$\mathrm{Sp}\,[C(x_0)] = \{h \in X : y_h(\omega)\lambda_0(\omega) = 0 \text{ 对几乎处处的 } \omega \in \Omega \text{ 成立}\}, \tag{5.85}$$

其中 $y_h := DG(x_0)h$.

证明　对 $y \in Y$, 考虑集合

$$\Delta(y) := \{\omega \in \Omega : y(\omega) = 0\},$$

它的定义是在忽视测度为零的集合意义下的. 令 $y_0 := G(x_0)$, 有

$$T_K(y_0) = \{y \in L_2(\Omega) : y(\omega) \leqslant 0 \text{ 对几乎处处的 } \omega \in \Delta(y_0) \text{ 成立}\}.$$

因为 $\langle\lambda_0, y_0\rangle = 0$, 得到 $\lambda_0(\omega) = 0$ 对几乎处处的 $\omega \in \Omega \setminus \Delta(G(x_0))$ 成立. 所以若 h 是临界方向, 则

$$\langle\lambda_0, DG(x_0)h\rangle = \int_{\Delta(y_0)} y_h(\omega)\lambda_0(\omega)\mathrm{d}\mu(\omega).$$

因为 $\lambda_0(\omega) \geqslant 0$ 对几乎处处的 $\omega \in \Omega$ 成立, 对临界方向 h, $y_h \in T_K(y_0)$ 且 $y_h(\omega) \leqslant 0$ 对几乎处处的 $\omega \in \Delta(y_0)$ 成立, 所以上述积分是零的充分必要条件是 $y_h(\omega)\lambda_0(\omega) = 0$ 对几乎处处的 $\omega \in \Delta(y_0)$ 成立. 于是

$$\begin{aligned} C(x_0) = \{h \in X : & \, y_h(\omega) \leqslant 0 \text{ 对几乎处处的 } \omega \in \Delta(y_0) \text{ 成立} \\ & y_h(\omega) = 0 \text{ 对几乎处处的 } \omega \in \Omega \setminus \Delta(y_0) \text{ 成立}\}. \end{aligned}$$

所以 (5.85) 右端包含临界锥, 因此也包含 Sp $[C(x_0)]$. 下证相反的包含关系亦成立. 给定 h 属于 (5.85) 的右端, 因为 $DG(x_0)$ 是映上的, 存在 $h_1 \in X$ 满足 $DG(x_0)h_1 = -(y_h)_+$(y_h 正部的负值), $DG(x_0)(h + h_1) = -(-y_h)_+$, 所以 $h = (h + h_1) - h_1$ 是两个临界方向的差, 证得 (5.85). 因为

$$[0, \lambda_0]_{K^-} = \{\mu \in L_2(\Omega) : 0 \leqslant \mu(\omega) \leqslant \lambda_0(\omega) \text{ 对几乎处处的 } \omega \in \Omega \text{ 成立}\},$$

所以等式 (5.71) 成立. 结论得证. □

推论 5.41 *设 $Y := L_2(\Omega), K := [L_2(\Omega)]_-$, x_0 与 λ_0 分别是相应的最优化问题 (5.48) 的局部最优解与 Lagrange 乘子. 设空间 X 是自反的, $DG(x_0)$ 是映上的, $Q(h) := D^2_{xx}L(x_0, \lambda_0)(h, h)$ 是 Legendre 形式, 则临界点(x_0, λ_0) 是强正则的充分必要条件是下述条件成立*

$$D^2_{xx}L(x_0, \lambda_0)(h, h) > 0, \quad \forall h \neq 0, \quad [DG(x_0)h](\omega)\lambda_0(\omega) = 0 \text{ 对几乎处处的 } \omega \in \Omega \text{ 成立}. \tag{5.86}$$

由定理 5.30 及引理 5.40 即得此结论.

3. *度量投影*

设 X 是 Hilbert 空间, 考虑最优化问题

$$\min_{x \in X} \|\psi(u) - x\| \quad \text{s.t.} \quad G(x, u) \in K, \tag{5.87}$$

以 $u \in U$ 为参数. 设 U 与 Y 是 Banach 空间, K 是 Y 的闭凸子集, 映射 $\psi : U \to X$, $G : X \times U \to Y$ 是二次连续可微的. 对固定的 $u \in U$, 上述问题的最优解 $\bar{x}(u)$ 可视为向量 $\psi(u)$ 到集合 $\Phi(u) := \{x \in X : G(x, u) \in K\}$ 上的度量投影, 即 $\bar{x}(u) = P_{\Phi(u)}(\psi(u))$.

定理 5.42　设 X 是 Hilbert 空间, $\psi(u)$ 与 $G(x,u)$ 是二次连续可微的, $\psi(u_0) \in \Phi(u_0)$, 映射 $G(\cdot, u_0)$ 与集合 K 在 $x_0 = \psi(u_0)$ 处的 Robinson 约束规范成立, 则存在 u_0 的邻域 $\mathcal{V}_U$ 满足下述结论:

(i) 对所有的 $u \in \mathcal{V}_U$, 最优化问题 (5.87) 具有唯一的最优解 $\bar{x}(u)$.

(ii) 最优解 $\bar{x}(u)$ 在 u_0 处是 Lipschitz 稳定的, 即

$$\|\bar{x}(u) - x_0\| = \mathrm{O}(\|u - u_0\|).$$

(iii) 存在某一 $\kappa > 0$,

$$\|\bar{x}(u) - \bar{x}(u')\| \leqslant \kappa \|u - u'\|^{\frac{1}{2}}, \quad \forall u, u' \in \mathcal{V}_U. \tag{5.88}$$

若进一步, $D_x G(x_0, u_0)$ 是映上的, 则 $\bar{x}(u)$ 在 u_0 的邻域内是 Lipschitz 连续的.

(iv) $\bar{x}(u)$ 在 $u = u_0$ 处是方向可微的且其方向导数$\bar{x}'(u_0, d)$ 是下述问题的最优解

$$\min_{h \in X} \|D\psi(x_0)d - h\| \quad \text{s.t.} \quad DG(x_0, u_0)(h, d) \in T_K(G(x_0, u_0)). \tag{5.89}$$

证明　设 $\bar{x}(u)$ 是 (5.87) 的最优解. 由稳定性定理, 对 u_0 的邻域中的所有 u, 存在可行点 $\hat{x}(u) \in \Phi(u)$ 满足 $\|\hat{x}(u) - x_0\| = \mathrm{O}(\|u - u_0\|)$. 进一步,

$$\begin{aligned} \|\bar{x}(u) - x_0\| &\leqslant \|\psi(u) - \bar{x}(u)\| + \|\psi(u) - x_0\| \\ &\leqslant \|\psi(u) - \hat{x}(u)\| + \|\psi(u) - x_0\| \\ &\leqslant \|\hat{x}(u) - x_0\| + 2\|\psi(u) - \psi(u_0)\|. \end{aligned}$$

可见, 只要证明了投影的存在性, 结论 (ii) 即得证.

显然, 可以把 (5.87) 的目标函数用函数 $f(x,u) := \|\psi(u) - x\|^2$ 来替换. 因为 X 是 Hilbert 空间, 函数 $f(x,u)$ 是二阶连续可微的且 $D_{xx}^2 f(x,u)(h,h) = 2\|h\|^2$. 由 $\psi(u_0) \in \Phi(u_0)$, 有非扰动问题 (即 $u = u_0$)(5.87) 具有唯一解 $x_0 = \psi(u_0)$. 进一步, 此种情况 $D_x f(x_0, u_0) = 0$, 从而非扰动问题具有 Lagrange 乘子 $\lambda_0 = 0$. 结果, 在 (x_0, u_0) 处的严格约束规范(见定义 4.46) 与所设的 Robinson 约束规范重合. 由命题 4.47, 对所有充分接近 u_0 的 $u \in U$, 问题 (5.87) 的 Larange 乘子一致地接近于 $0 \in Y^*$. 由连续性假设可得, 对 $(x_0, 0, u_0)$ 的某一邻域中所有的 (x, λ, u), $D_{xx}^2 L(x, \lambda, u)(h,h) \geqslant \|h\|^2$. 由注记 5.32 所讨论的定理 5.27 的变形可推出, 相对于所考虑的参数化, 一致二阶增长条件在 x_0 处成立.

注意到这些参数化包含倾斜参数化. 定理 5.17 可推出结论 (i) 与 (iii). 还注意到, 由 $\bar{x}(u)$ 在 $u = u_0$ 处的连续性 (由 (ii) 可得), 所以最优解 $\bar{x}(u)$ 是全局唯一的, 即 $\bar{x}(u)$ 不仅仅是 x_0 的邻域上的, 而且是空间 X 上的唯一最优解.

还需证明 (iv). 为此可应用 4.7.2 节的结果. 因为非扰动问题唯一的 Lagrange 乘子是 0, 子问题 $(\mathcal{D}\mathcal{Q}^R)$ (这里简化为 (5.89)) 与 4.7.1 节中的问题 $(\mathcal{D}\mathcal{Q})$ 重合. 结合定理 4.85 与 4.91, 有 (4.235) 成立. 结论由定理 4.95 得到. □

例 4.99 表明, 在 Robinson 约束规范下, (5.88) 界的指数常数 $\frac{1}{2}$ 是不能改进的.

现在考虑 (固定的) 闭集合$\Phi \subset X$, 度量投影集值映射 $\Pi_\Phi(x)$, 即 $\Pi_\Phi(x)$ 是 Φ 中距离 x 最近的点的集合. 若集合 Φ 是凸的, 则 $\Pi_\Phi(x) = \{P_\Phi(x)\}$ 是单点集对每一 $x \in X$ 均成立, 即度量投影集值映射 $P_\Phi(\cdot)$ 是有定义的. 由上面的定理还会得到, 若 $\Phi := G^{-1}(K)$, $G : X \to Y$ 是二次连续可微的, 则集合 $K \subset Y$ 是闭凸集, 若 Robinson 约束规范 在 $x_0 \in \Phi$ 处成立, 则对 x_0 的某一邻域中的所有 x, $P_\Phi(x)$ 存在唯一且是 Lipschitz 连续的, 进一步, $P_\Phi(x)$ 在 x_0 处是方向可微的且方向导数$P'_\Phi(x_0, h)$ 由 h 到 $T_\Phi(x_0)$ 的投影得到, 即

$$P'_\Phi(x_0, h) = P_{T_\Phi(x_0)}(h). \tag{5.90}$$

即使 Φ 是凸集, 映射 $P_\Phi(\cdot)$ 在点 $x_0 \notin \Phi$ 处亦可能不是方向可微的, 这样的例子在二维空间 $X := \mathbb{R}^2$ 中已存在. 例如, 考虑由例 3.31 中构造的凸集合 $S \subset \mathbb{R}^2$. 通过直接计算可以证明, 相应的度量投影映射 $P_S(\cdot)$ 在点 $x_0 := (0, -1)$ 沿方向 $h := (1, 0)$ 不是方向可微的. 由第 4 章的一般性理论, 若 Φ 是凸集, 或者 Φ 是多面体集合, 或者 X 是有限维的且 Φ 是二阶正则的, 则 $P_\Phi(\cdot)$ 是方向可微的.

5.2 非线性规划

这一节的目的在于给出非线性规划问题, 即具有有限多个等式与不等式约束的有限维最优化问题的一个基础性的讨论.

其他的章节已经包含了大量处理非线性规划的结果, 这些结果通过作为一般性最优化问题结果的特例而得到. 这一节讨论这些结果中的一些是通过 $\mathbb{R}^n$ 中的标准微分分析与有限维线性规划的理论得到的. 这一节自成体系, 仅仅用到其他章节的少量结果, 其证明也是通过初等的工具得到的. 除非特别声明, 设 $x \cdot y = \sum_{i=1}^n x_i y_i$ 记两个向量 $x, y \in \mathbb{R}^n$ 的标准的数乘, $\|x\| := (x^{\mathrm{T}}x)^{\frac{1}{2}}$ 记向量 $x \in \mathbb{R}^n$ 的欧氏范数. 处理强稳定性与强正则性的具体结果在 5.1.6 节中已经得到.

5.2.1 有限维的线性规划

这一节讨论有限维线性规划问题, 即在有限维空间中目标函数是线性的, 具有有限多个线性约束的最优化问题. 在线性规划的理论中, 下述结果即是著名的 Farkas 引理.

引理 5.43 (Farkas 引理) 对任何向量 $a_i \in \mathbb{R}^n, i = 0, \cdots, p$, 下述两个条件是等价的:

(i) 对满足 $a_i \cdot h = 0, i = 1, \cdots, q$, $a_i \cdot h \leqslant 0, i = q+1, \cdots, p$ 的 $h \in \mathbb{R}^n$, 不等式 $a_0 \cdot h \geqslant 0$ 成立.

(ii) 存在 $\lambda \in \mathbb{R}^p$ 满足 $a_0 + \sum_{i=1}^{p} \lambda_i a_i = 0$ 且 $\lambda_i \geqslant 0$ 对 $i = q+1, \cdots, p$ 成立.

证明 把等式 $a_i \cdot h = 0$ 表示为两个不等式 $a_i \cdot h \leqslant 0$ 与 $-a_i \cdot h \leqslant 0$, 在上述引理中可以将所有的等式替换为相应的线性不等式. 因此, 不失一般性, 设系统中只有不等式出现.

(ii)⇒(i) 是显然的. 事实上, 若 h 满足 $a_i \cdot h \leqslant 0, 1, \cdots, p$, 则由于 $\lambda \in \mathbb{R}^p_+$, 有 $a_0 \cdot h = -\sum_{i=1}^{p} \lambda_i a_i \cdot h \geqslant 0$.

现在证 (i) 推出 (ii). 由命题 2.41, 凸集合

$$E := \left\{ \sum_{i=1}^{p} \lambda_i a_i : \lambda_i \geqslant 0, i = 1, \cdots, p \right\}$$

是闭的. 需要证明 $-a_0 \in E$. 设这一包含关系不成立, 令 $\bar{h}$ 是 $-a_0$ 到 E 上的度量 (直交) 投影, 即 $\bar{h}$ 是下述最优化问题的最优解

$$\min_{h \in E} \quad \|a_0 + h\|^2.$$

注意到, 因为 E 是非空的凸的闭集合, 目标函数是连续的强凸的 (引理 2.33), 此问题有唯一的最优解. 令 $b := a_0 + \bar{h}$. 由 $-a_0 \notin E$, 有 $b \neq 0$. 因为 E 是凸的, 所以对任何 $e \in E$, $t \in [0,1]$, $\bar{h} + t(e - \bar{h}) \in E$, 有

$$b \cdot (e - \bar{h}) = \frac{1}{2} \lim_{t \downarrow 0} t^{-1}(\|a_0 + \bar{h} + t(e - \bar{h})\|^2 - \|a_0 + \bar{h}\|^2) \geqslant 0.$$

取 $e := 0$, 得 $b \cdot \bar{h} \leqslant 0$, 因此 $b \cdot a_0 = b \cdot (b - \bar{h}) > 0$. 另一方面, 取 $e := t^{-1} a_i$, 其中 $i \in \{1, \cdots, p\}$, 有

$$0 \leqslant \lim_{t \downarrow 0} t(t^{-1} a_i - \bar{h}) \cdot b = b \cdot \bar{a}_i.$$

在条件 (i) 中取 $h := -b$, 得到矛盾. 证毕. □

考虑下述的线性规划问题

$$\text{(LP)} \qquad \begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c \cdot x, \\ \text{s. t.} \quad & a_i \cdot x = b_i, \quad i = 1, \cdots, q, \\ & a_i \cdot x \leqslant b_i, \quad i = q+1, \cdots, p. \end{aligned} \tag{5.91}$$

这一问题的对偶问题可表示为

$$\text{(LD)}\qquad \begin{aligned} \max_{x\in\mathbb{R}^p}\quad & -b\cdot\lambda, \\ \text{s.t.}\quad & c+\sum_{i=1}^{p}\lambda_i a_i=0, \\ & \lambda_i\geqslant 0,\quad i=q+1,\cdots,p. \end{aligned} \tag{5.92}$$

注意到, 原始问题 (LP) 可以表示为极小–极大问题

$$\min_{x\in\mathbb{R}^n}\Big\{\sup_{\lambda\in\mathbb{R}^p} L(x,\lambda):\lambda_i\geqslant 0, i=q+1,\cdots,p\Big\}, \tag{5.93}$$

其中

$$L(x,\lambda):=c\cdot x+\sum_{i=1}^{p}\lambda_i(a_i\cdot x-b_i)$$

是相应的 Lagrange 函数. (5.93) 通过交换极大运算与极小运算可以得到相应的对偶问题(LD). 所以 (LD) 可视为 (LP) 的 Lagrange 对偶. 注意到, 对偶问题(LD) 可以表示为原始问题 (LP) 的形式. 写出 (LD) 的对偶即得到原始问题 (LP). 所以, 原始问题与对偶问题间存在完全的对称, 哪一个称为原始的, 哪一个是对偶的, 可以任意选定.

定理 5.44 (i) 点 $x\in\mathbb{R}^n$ 与 $\lambda\in\mathbb{R}^p$ 分别是 (LP) 与 (LD) 的最优解的充分必要条件是这些点是可行的且下述互补条件成立:

$$\lambda_i(a_i\cdot x-b_i)=0,\quad i=q+1,\cdots,p. \tag{5.94}$$

(ii) 问题 (LP) 与 (LD) 具有相同的最优值, 除非 (LP) 与 (LD) 都是不相容的. 若它们 (公共的) 最优值是有限的, 则两个问题均有非空的最优解集.

证明 令 x 与 λ 是问题 (LP) 与 (LD) 的可行点. 因为 λ 是 (LD) 的可行点, 目标函数值之差可以表示为

$$c\cdot x+b\cdot\lambda=-\left(\sum_{i=1}^{p}\lambda_i a_i\right)\cdot x+b\cdot\lambda=-\sum_{i=1}^{p}\lambda_i(a_i\cdot x-b_i). \tag{5.95}$$

因为 x 是可行的, 因此满足 (LP) 的约束, 由 λ 的后一部分分量的非负性得到, (5.95) 右端的表示是非负的, 因此它等于零当且仅当互补条件成立. 我们看到, 对任何相应问题的可行点 x 与 λ, 原始问题的相应值 $c\cdot x$ 大于或等于对偶问题的值 $-b\cdot\lambda$, 等式成立当且仅当互补条件成立. 有 val(LP)$\geqslant$ val(LD), 若 x 与 λ 是可行的且互补条件成立, 则 x 与 λ 是最优的.

如果 (LP) 与 (LD) 均是不相容的, 上述结论显然成立. 所以, 可设这两个问题之一, 如 (LP) 是相容的, 因此 val(LP)$< +\infty$. 因为 val(LP)$\geqslant$ val(LD), 所以如果 val(LP)$=-\infty$, 则 val(LD)$=-\infty$, 因此, 此种情况结论成立. 所以可设 (LP) 具有有限的最优值. 由定理 2.198 得 (LP) 至少有一最优解 $\bar{x}$, 记

$$I(x) := \{i : a_i \cdot x = b_i,\ i = q+1, \cdots, p\}$$

为在 x 处起作用的不等式约束的指标集, 若 $h \in \mathbb{R}^n$ 满足

$$a_i \cdot h = 0, \quad i = 1, \cdots, q; \quad a_i \cdot h \leqslant 0, \quad i \in I(\bar{x}),$$

则对充分小的 $t > 0$, 有 $\bar{x} + th$ 是 (LP) 的可行点. 因为 $\bar{x}$ 是 (LP) 的最优解, 可得到 $c \cdot \bar{x} \leqslant c \cdot (\bar{x} + th)$, 因此 $c \cdot h \geqslant 0$. 则由 Farkas 引理(引理 5.43), 存在 $\lambda \in \mathbb{R}^p$ 满足

$$c + \sum_{i=1}^{p} \lambda_i a_i = 0; \quad \lambda_i \geqslant 0, \quad q+1 \leqslant i \leqslant p; \quad \lambda_i = 0, \quad i \in \{q+1, \cdots, p\} \setminus I(\bar{x}).$$

于是得 λ 是 (LD) 的可行点, 互补条件成立. 由 (5.95) 可得 λ 是 (LD) 的最优解, 且 val(LP)= val(LD). 因为问题 (LP) 与 (LD) 间没有对偶间隙, 由 (5.95) 可得, 若 x 与 λ 是最优的, 则互补条件成立. 证毕. □

现在讨论线性规划问题的原始最优解集或对偶最优解集的有界性条件. 这一结果将用在非线性规划问题约束规范的分析中. 考虑对偶的线性规划问题对 (LP) 与 (LD).

命题 5.45　下述两个条件是等价的:

(i) 向量 $a_i, i = 1, \cdots, q$ 是线性无关的, 存在 $x \in \mathbb{R}^n$, 满足

$$a_i \cdot x = b_i, \quad i = 1, \cdots, q; \quad a_i \cdot x < b_i, \quad i = q+1, \cdots, p. \tag{5.96}$$

(ii) 齐次系统

$$b \cdot \lambda \leqslant 0, \quad \sum_{i=1}^{p} \lambda_i a_i = 0, \quad \lambda_i \geqslant 0, \quad i = q+1, \cdots, p \tag{5.97}$$

仅有一解 $\overline{\lambda} = 0$.

若还有, (LP) 有有限的最优值, 则上述条件 (i) 与 (ii) 等价于下述条件:

(iii) 下述齐次系统具有唯一的 (零) 解

$$b \cdot \lambda = 0, \quad \sum_{i=1}^{p} \lambda_i a_i = 0, \quad \lambda_i \geqslant 0, \quad i = q+1, \cdots, p. \tag{5.98}$$

(iv) (LD) 的最优解集 $\mathcal{S}$(LD) 是非空有界的.

证明 先证 (i) 可推出 (ii). 设 x 满足 (5.96), 考虑

$$\varepsilon := \min\{b_i - a_i \cdot x : i = q+1, \cdots, p\},$$

有 $\varepsilon > 0$, 若 λ 是 (5.97) 的解, 则

$$0 = \left(\sum_{i=1}^{p} \lambda_i a_i\right) \cdot x = \sum_{i=1}^{p} \lambda_i a_i \cdot x \leqslant b \cdot \lambda - \varepsilon \sum_{i=q+1}^{p} \lambda_i \leqslant 0.$$

可推出 $b \cdot \lambda = 0$ 且 $\lambda_i = 0,\ i = q+1, \cdots, p$, 因此 $\sum\limits_{i=1}^{q} \lambda_i a_i = 0$. 因为向量 $a_i,\ i = 1, \cdots, q$ 是线性无关的, 所以 $\lambda = 0$.

相反地, 设 (ii) 成立而 (i) 不成立. 若向量 $a_i,\ i = 1, \cdots, q$ 不是线性无关的, 则存在非零的 $\lambda \in \mathbb{R}^p$ 满足 $\lambda_i = 0$, 对所有的 $i = q+1, \cdots, p$, 且 $\sum\limits_{i=1}^{q} \lambda_i a_i = 0$. 显然, 向量 λ 满足 (5.97), 因此与 (ii) 相矛盾. 需要讨论的另外的情况是向量 $a_i,\ i = 1, \cdots, q$ 线性无关, 但 (5.96) 没有解的情况. 这意味着线性规划

$$\begin{aligned}
\min_{(x,z) \in \mathbb{R}^n \times \mathbb{R}} \quad & z, \\
\text{s.t.} \quad & a_i \cdot x = b_i, \qquad i = i = 1, \cdots, q, \\
& a_i \cdot x \leqslant b_i + z, \quad i = q+1, \cdots, p
\end{aligned}$$

有非负的最优值. 以下验证这一线性规划是相容的. 因为向量 $a_i,\ i = 1, \cdots, q$ 是线性无关的, 存在 x_0 满足等式约束. 可以取 z 充分大, 使不等式约束也成立. 所以相应的对偶问题

$$\max_{\lambda} \ -b \cdot \lambda \quad \text{s.t.} \quad \sum_{i=1}^{p} \lambda_i a_i = 0, \quad \lambda_i \geqslant 0, \quad i = q+1, \cdots, p, \quad \sum_{i=q+1}^{p} \lambda_i = 1$$

有非负的最优值和至少一最优解. 这一最优解导致与 (ii) 的矛盾.

现在设 (LP) 有最优解, 因此 (LD) 也有最优解. 显然, (ii) 可推出 (iii). 相反地, 假设 (ii) 不成立, 即存在 $\lambda \neq 0$ 满足 (5.98). 若 $b \cdot \lambda = 0$, 则显然 (iii) 不成立, 否则, $b \cdot \lambda < 0$. 此种情况, 对给定的某 $\bar{\lambda} \in \mathcal{S}$(LD), $\bar{\lambda} + \lambda$ 是 (LD) 的可行点, 相应的目标值 $-b \cdot (\bar{\lambda} + \lambda)$ 比最优值 $-b \cdot \bar{\lambda} =$val(LD) 严格大, 导致矛盾. 证得 (ii) 与 (iii) 的等价性.

因为 (LD) 有最优解, $\mathcal{S}$(LD) 的回收锥是 (LD) 的可行集的回收锥与同 b 垂直的超平面的交集, 因此等于满足 (5.97) 的 λ 的集合. 因为有限维空间中的非空凸子

集是有界的充分必要条件是其回收锥是 $\{0\}$(见 2.1.4 节中的讨论), 所以 $\mathcal{S}(\text{LD})$ 是有界的充分必要条件是 (5.98) 只有平凡的解 $\lambda = 0$, 证得 (iii) 与 (iv) 的等价性. □

因为在上述的线性规划框架下, 原始问题与对偶问题是彼此对称的, 上述结果还给出了原始问题的最优解集有界性的一个刻画.

5.2.2 非线性规划的最优性条件

这一节的其余部分考虑如下形式的非线性规划问题:

$$
\text{(NLP)} \qquad \begin{aligned} \min \quad & f(x), \\ \text{s. t.} \quad & g_i(x) = 0, \quad i = 1, \cdots, q, \\ & g_i(x) \leqslant 0, \quad i = q+1, \cdots, p. \end{aligned} \tag{5.99}
$$

设目标函数 $f : \mathbb{R}^n \to \mathbb{R}$ 与约束映射 $G := (g_1, \cdots, g_p) : \mathbb{R}^n \to \mathbb{R}^p$ 是二次连续可微的. 讨论上述问题 (NLP) 的一阶与二阶最优性条件.

用 $I(x)$ 记在 x 处起作用的不等式约束集合

$$
I(x) := \{i : g_i(x) = 0, i = q+1, \cdots, p\}.
$$

很自然, 把 (NLP) 的可行点 $\bar{x}$ 与目标函数、等式约束和在 $\bar{x}$ 处起作用的不等式约束的线性化得到的问题相联系, 即

$$
\begin{aligned} \min_{h \in \mathbb{R}^n} \quad & Df(\bar{x})h, \\ \text{s.t.} \quad & Dg_i(\bar{x})h = 0, \quad i = 1, \cdots, q, \\ & Dg_i(\bar{x})h \leqslant 0, \quad i \in I(\bar{x}). \end{aligned} \tag{5.100}
$$

基于下述的简单观察, 还有另一种线性化. 若点 $\bar{x}$ 是 (NLP) 的局部最优解, 则 $(\bar{x}, 0)$ 是下述非线性规划问题的局部极小点:

$$
\begin{aligned} \min_{(x,z) \in \mathbb{R}^n \times \mathbb{R}} \quad & z, \\ \text{s.t.} \quad & g_i(x) = 0, \quad i = 1, \cdots, q, \\ & g_i(x) \leqslant z, \quad i = q+1, \cdots, p, \\ & f(x) - f(\bar{x}) \leqslant z. \end{aligned} \tag{5.101}
$$

在 $(\bar{x}, 0)$ 处相应的线性化问题是

$$
\begin{aligned} \min_{(h,z) \in \mathbb{R}^n \times \mathbb{R}} \quad & z, \\ \text{s.t.} \quad & Dg_i(\bar{x})h = 0, \quad i = 1, \cdots, q, \\ & Dg_i(\bar{x})h \leqslant z, \quad i \in I(\bar{x}), \\ & Df(\bar{x})h \leqslant z. \end{aligned} \tag{5.102}
$$

与问题 (NLP) 相联系的 Lagrange 函数与广义 Lagrange 函数分别定义为

$$L(x,\lambda):=f(x)+\sum_{i=1}^{p}\lambda_i g_i(x),\quad L^g(x,\lambda_0,\lambda):=\lambda_0 f(x)+\sum_{i=1}^{p}\lambda_i g_i(x),$$

其中 $\lambda_0\in\mathbb{R},\lambda\in\mathbb{R}^p$.

定义 5.46 问题 (NLP) 的可行点 $\bar{x}$ 处的广义 Lagrange 乘子集合 $\Lambda^g(\bar{x})$ 定义为满足下述一阶最优性条件的非零向量 $(\lambda_0,\lambda)=(\lambda_0,\lambda_1,\cdots,\lambda_p)$ 的全体：

$$D_xL^g(\bar{x},\lambda_0,\lambda)=0,\quad \lambda_0\geqslant 0,\quad \lambda_i\geqslant 0,\quad \lambda_i g_i(\bar{x})=0,\quad i=q+1,\cdots,p.$$

若广义 Lagrange 乘子(λ_0,λ) 满足 $\lambda_0=0$, 则称 λ 是奇异 Lagrange 乘子. 若 $\lambda_0=1$, 称 λ 是 Lagrange 乘子,Lagrange 乘子集合记为 $\Lambda(\bar{x})$.

不难验证 (见命题 3.14), 若 Lagrange 乘子集合 $\Lambda(\bar{x})$ 是非空的, 则奇异 Lagrange 乘子连同 0 构成的集合构成了 $\Lambda(\bar{x})$ 的回收锥.

令 $\bar{x}$ 是 (NLP) 的可行点, 称 Mangasarian-Fromovitz(MF) 约束规范在 $\bar{x}$ 处成立, 若下述条件被满足：

(i) 向量 $Dg_i(\bar{x}),\ i=1,\cdots,q$ 是线性无关的.

(ii) $\exists\bar{h}\in\mathbb{R}^n: Dg_i(\bar{x})\bar{h}=0, i=1,\cdots,q;\ Dg_i(\bar{x})\bar{h}<0, i\in I(\bar{x})$.

命题 5.47 设 $\bar{x}$ 是 (NLP) 的局部最优解, 则广义 Lagrange 乘子集合 $\Lambda^g(\bar{x})$ 是非空的, 且下述条件是等价的：

(i) Mangasarian-Fromovitz 约束规范在 $\bar{x}$ 处成立.

(ii) 奇异 Lagrange 乘子集合是空集.

(iii) Lagrange 乘子集合 $\Lambda(\bar{x})$ 非空有界.

证明 先考虑向量 $Dg_i(\bar{x}), i=1,\cdots,q$ 不是线性无关的情况, 则存在非零的向量 $\mu\in\mathbb{R}^p$ 满足 $\sum_{i=1}^{q}\mu_i Dg_i(\bar{x})=0$. 置 $\lambda_i:=\mu_i,\ i=1,\cdots,q,\ \lambda_i:=0,\ i=q+1,\cdots,p$, 得到 λ 是奇异 Lagrange 乘子.

现在设向量 $Dg_i(\bar{x}),\ i=1,\cdots,q$ 是线性无关的. 我们给出论断：线性化问题 (5.102) 的最优值是零. 事实上, 如果不是这样, 则存在 $h\in\mathbb{R}^n$ 满足 $Dg_i(\bar{x})h=0,\ i=1,\cdots,q,\ Df(\bar{x})h<0, Dg_i(\bar{x})h<0, i\in I(\bar{x})$. 由于假设 $Dg_i(\bar{x}),\ i=1,\cdots,q$ 是线性无关的, 由隐函数定理, 存在路径 $x(t)=\bar{x}+th+o(t), t\geqslant 0$ 满足 $g_i(x(t))=0, i=1,\cdots,q$ 对充分小的 $t>0$ 成立. 对每一 $i\in I(\bar{x})$, 有 $g_i(x(t))=tDg_i(\bar{x})h+o(t)<0$, 因此 $x(t)$ 是可行的. 同时, $f(x(t))=f(\bar{x})+tDf(\bar{x})h+o(t)<f(\bar{x})$, 这与 $\bar{x}$ 的局部最优性是矛盾的.

问题 (5.102) 是线性规划问题, 其对偶问题为

$$\begin{aligned} &\max_{(\lambda_0,\lambda)} \quad 0, \\ &\text{s.t.} \quad D_x L^g(\bar{x},\lambda_0,\lambda)=0,\ \lambda_0+\sum_{i\in I(\bar{x})}\lambda_i=1,\ \lambda_i\geqslant 0, i\in I(\bar{x})\cup\{0\}. \end{aligned}$$

由于问题 (5.102) 的最优值是零, 由定理 5.44 得到, 其对偶问题的最优值也是零. 结果, 对偶问题的可行集是非空的. 不难看到, 上述对偶问题的任何可行点 (λ_0,λ) 均是 (NLP) 问题与 $\bar{x}$ 相联系的广义 Lagrange 乘子, 这表明集合 $\Lambda^g(\bar{x})$ 是非空的.

现在证明, 由 MF 约束规范可得线性化问题(5.100) 的最优值是零. 事实上, 令 h 是 (5.100) 的可行点, $\bar{h}$ 是满足 MF 约束规范条件 (ii) 的向量. 对给定的 $\varepsilon>0$, 考虑 $h_\varepsilon:=h+\varepsilon\bar{h}$. 因为对所有 $i\in I(\bar{x})$ 及充分小的 $\varepsilon>0$, 有 $Dg_i(\bar{x})h_\varepsilon<0$, 类似于本证明开始时用到的推证, 存在路径 $x(t)=\bar{x}+th_\varepsilon+o(t)$ 满足 $g_i(x(t))=0,\ i=1,\cdots,q$, $g_i(x(t))\leqslant 0,\ i=q+1,\cdots,p$ 对 $t>0$ 充分小时成立, 因而对 $t>0$ 充分小, 该路径是可行的. 由于 $\bar{x}$ 是局部极小点, 有

$$Df(\bar{x})h_\varepsilon=\lim_{t\downarrow 0}\frac{f(\bar{x}+th_\varepsilon)-f(\bar{x})}{t}\geqslant 0.$$

由于对任何充分小的 $\varepsilon>0$ 均是成立的, 所以对 (5.100) 的任何可行点 h 均有 $Df(\bar{x})h\geqslant 0$. 证得 $h=0$ 是 (5.100) 的最优解.

(5.100) 是线性规划问题, 其对偶问题可以表示为下述形式:

$$\begin{aligned} &\max_{\lambda} \quad 0 \\ &\text{s. t.} \quad D_x L(\bar{x},\lambda)=0,\quad \lambda_i\geqslant 0,\quad i\in I(\bar{x}). \end{aligned}$$

若 (5.100) 的最优值是零, 由定理 5.44 可得, 对偶问题的最优值取零, 则对偶问题的可行集是非空的. 于是 MF 约束规范可推出 Lagrange 乘子的存在性.

最后, 应用命题 5.45, 其中 x 对应 h, a_i 对应 $g_i(x)$, b 是零. 这一命题的关系 (i) 恰好是 MF 约束规范, 而 (5.97) 刻画了广义 Lagrange 乘子(排除 $\lambda=0$). 因此, 这一命题 (i) 与 (ii) 的等价性意味着 MF 约束规范成立的充分必要条件是奇异 Lagrange 乘子集合是空集. 若后者成立, 由于线性化问题 (5.100) 具有最优值 0, 由命题 5.45, Lagrange 乘子集合是非空有界的. 相反地, 设 Lagrange 乘子集合是非空有界的, 则相应的线性规划 (5.100) 具有最优值 0, 再次由命题 5.45 得 MF 约束规范成立. 证毕. □

现在讨论二阶最优性条件. 与 (NLP) 的可行点 $\bar{x}$ 相联系的临界锥可表示为下述形式:

$$C(\bar{x}):=\{h: Df(\bar{x})h\leqslant 0, Dg_i(\bar{x})h=0,\ i\leqslant q,\quad Dg_i(\bar{x})h\leqslant 0,\ i\in I(\bar{x})\}.$$

它的元素称为临界方向. 若 Lagrange 乘子集合 $\Lambda(\bar{x})$ 是非空的, 则临界锥 $C(\bar{x})$ 的定义中的不等式 $Df(\bar{x})h \leqslant 0$ 可替换为等式 $Df(\bar{x})h = 0$, 且对任何 $\overline{\lambda} \in \Lambda(\bar{x})$, 临界锥可表示为 (5.77) 的形式.

用 $I(\bar{x},h)$ 记在 $\bar{x}$ 处起作用的且沿方向 h 的一阶量也是起作用的约束集合, 即

$$I(\bar{x},h) := \{i \in I(\bar{x}) : Dg_i(\bar{x})h = 0\}.$$

命题 5.48 (二阶最优性条件) *设 $\bar{x}$ 是问题 (NLP) 的可行点. 则下述结论成立:*

(i) *若 $\bar{x}$ 是 (NLP) 的局部最优解, 则对每一 $h \in C(\bar{x})$, 存在广义 Lagrange 乘子 $(\lambda_0, \lambda) \in \Lambda^g(\bar{x})$ 满足*

$$D^2_{xx}L^g(\bar{x}, \lambda_0, \lambda)(h,h) \geqslant 0. \tag{5.103}$$

(ii) *若对每一 $h \in C(\bar{x}) \setminus \{0\}$, 存在 $(\lambda_0, \lambda) \in \Lambda^g(\bar{x})$ 满足*

$$D^2_{xx}L^g(\bar{x}, \lambda_0, \lambda)(h,h) > 0, \tag{5.104}$$

则 $\bar{x}$ 是 (NLP) 的满足二阶增长条件的局部最优解.

证明 (i) 令 $h \in C(\bar{x})$ 是临界方向. 首先考虑 $Dg_i(\bar{x})$, $i = 1, \cdots, q$ 不是线性无关的情况. 存在奇异 Lagrange 乘子 $\hat{\lambda}$ 满足 $\hat{\lambda}_i = 0$, $i = q+1, \cdots, p$. 若 $\lambda := \hat{\lambda}$, 则 (5.103) 式成立, 得到结论, 否则, 由于 $-\hat{\lambda}$ 是另一奇异 Lagrange 乘子, 且

$$D^2_{xx}L^g(\bar{x}, 0, -\hat{\lambda})(h,h) = -D^2_{xx}L^g(\bar{x}, 0, \hat{\lambda})(h,h),$$

则取 $\lambda := -\hat{\lambda}$ 时, (5.103) 式成立.

现在讨论 $Dg_i(\bar{x})$, $i = 1, \cdots, q$ 是线性无关的情形. 考虑线性规划

$$\begin{aligned} \min_{(w,z)\in\mathbb{R}^n\times\mathbb{R}} \quad & z, \\ \text{s. t.} \quad & Df(\bar{x})w + D^2 f(\bar{x})(h,h) \leqslant z, \\ & Dg_i(\bar{x})w + D^2 g_i(\bar{x})(h,h) = 0, \ i = 1, \cdots, q, \\ & Dg_i(\bar{x})w + D^2 g_i(\bar{x})(h,h) \leqslant z, \ i \in I(\bar{x},h). \end{aligned} \tag{5.105}$$

这一问题的最优值是 0. 否则, 存在 w 满足

$$\begin{aligned} & Df(\bar{x})w + D^2 f(\bar{x})(h,h) < 0, \\ & Dg_i(\bar{x})w + D^2 g_i(\bar{x})(h,h) = 0, \quad i = 1, \cdots, q, \\ & Dg_i(\bar{x})w + D^2 g_i(\bar{x})(h,h) < 0, \quad i \in I(\bar{x},h). \end{aligned} \tag{5.106}$$

因为 $\bar{x}$ 是可行点, h 是临界方向, $g_i(\bar{x}) = 0$, $Dg_i(\bar{x})h = 0, i = 1, \cdots, q$, 由 (5.106) 的第二方程, 得到

$$g_i\Big(\bar{x} + th + \frac{1}{2}t^2 w\Big) = \frac{1}{2}t^2[Dg_i(\bar{x})w + D^2 g_i(\bar{x})(h,h)] + o(t^2) = o(t^2).$$

由隐函数定理, 存在路径 $x(t)=\bar{x}+th+\frac{1}{2}t^2w+o(t^2)$ 满足 $g_i(x(t))=0, i=1,\cdots,q$ 对充分小的 $t>0$ 成立, 则由二阶 Taylor 展开式, 对充分小的 $t>0$ 有

$$f(x(t))=f(\bar{x})+tDf(\bar{x})h+\frac{1}{2}t^2[Df(\bar{x})w+D^2f(\bar{x})(h,h)]+o(t^2)<f(\bar{x}),$$

类似地, 对所有的 $i\in I(\bar{x},h)$, 有 $g_i(x(t))<0$. 若 $i>q$ 且 $i\notin I(\bar{x},h)$, 则或 $g_i(\bar{x})<0$ 或 $g_i(\bar{x})=0$ 且 $Dg_i(\bar{x})h<0$, 两种情形时, 对充分小的 $t>0$, 均有 $g_i(x(t))<0$. 因此, 对充分小的 $t>0$, $x(t)$ 是可行的且 $f(x(t))<f(\bar{x})$, 这与 $\bar{x}$ 的局部最优性相矛盾. 于是证得 (5.105) 有非负的最优值.

因为 $Dg_i(\bar{x}), i=1,\cdots,q$ 是线性无关的, (5.105) 中的等式约束有可行解, 因此, 由于 z 可取任意大, 问题 (5.105) 是相容的. 所以, (5.105) 具有有限的非负的最优值. 因为 (5.105) 是线性规划问题, 其对偶问题具有相同的最优值. 问题 (5.105) 的对偶问题是

$$\begin{aligned}\max_{\lambda\in\mathbb{R}^p}\quad & D^2_{xx}L^g(\bar{x},\lambda_0,\lambda)(h,h),\\ \text{s. t.}\quad & D_xL^g(\bar{x},\lambda_0,\lambda)=0,\ \ \lambda_0+\sum_{i\in I(\bar{x},h)}\lambda_i=1,\\ & \lambda_0\geqslant 0,\ \lambda_i\geqslant 0,\ i\in I(\bar{x},h), \lambda_i=0,\ i>q,\ i\notin I(\bar{x},h).\end{aligned}$$

因为这一对偶问题的最优解是与 $\bar{x}$ 相联系的广义 Lagrange 乘子, 其对偶目标函数是 $D^2_{xx}L^g(\bar{x},\lambda_0,\lambda)(h,h)$①, 结论 (i) 得证.

现在考虑结论 (ii). 假设命题 (ii) 的结论不真, 则存在可行点序列 $x_k\to\bar{x}$ 满足 $f(x_k)\leqslant f(\bar{x})+o(\|x_k-\bar{x}\|^2)$. 置 $t_k:=\|x_k-\bar{x}\|$, 则

$$\limsup_{k\to\infty}\frac{f(x_k)-f(\bar{x})}{t_k^2}\leqslant 0.$$

如有必要, 可抽取一子列, 不妨设 $h_k:=(x_k-\bar{x})/t_k$ 收敛到单位范数的向量 $\hat{h}$, 即 $x_k=\bar{x}+t_k\hat{h}+o(t_k), \|\hat{h}\|=1$. 因为 $f(x_k)\leqslant f(\bar{x})+o(\|x_k-\bar{x}\|^2)$ 且 x_k 是可行的, 所以由 $f(x_k)$, $g_i(x_k), i=1,\cdots,p$ 的一阶展式可得 $\hat{h}$ 是临界方向. 令 $(\hat{\lambda}_0,\hat{\lambda})$ 是广义 Lagrange 乘子满足

$$\alpha:=D^2_{xx}L^g(\bar{x},\hat{\lambda}_0,\hat{\lambda})(\hat{h},\hat{h})>0.$$

则由于 $\hat{\lambda}$ 的对应于不等式约束的分量是非负的, $D_xL^g(\bar{x},\hat{\lambda}_0,\hat{\lambda})=0$, 有

$$\begin{aligned}\hat{\lambda}_0f(x_k)\geqslant L^g(x_k,\hat{\lambda}_0,\hat{\lambda})&=\hat{\lambda}_0f(\bar{x})+\frac{1}{2}t_k^2D^2_{xx}L^g(\bar{x},\hat{\lambda}_0,\hat{\lambda})(\hat{h},\hat{h})+o(t_k^2)\\ &\geqslant\hat{\lambda}_0f(\bar{x})+\frac{1}{2}\alpha t_k^2+o(t_k^2).\end{aligned}$$

① 原著中为 $D^2_{xx}L^g(\bar{x},\lambda)(h,h)$.

可得到

$$\alpha \leqslant 2\hat{\lambda}_0 \limsup_{k\to\infty} \frac{f(x_k)-f(\bar{x})}{t_k^2} \leqslant 0,$$

导致矛盾. □

注 5.49 若 Lagrange 乘子集合 $\Lambda(\bar{x})$ 是非空的, 则命题 5.48(ii) 的二阶充分性条件等价于下述 (更强的) 条件: 对任何 $h \in C(\bar{x}) \setminus \{0\}$, 存在 $\lambda \in \Lambda(\bar{x})$, 满足 $D_{xx}^2 L(\bar{x},\lambda)(h,h) > 0$. 这些条件也可以写为下述的等价形式:

$$\sup_{\lambda\in\Lambda(\bar{x})} D_{xx}^2 L(\bar{x},\lambda)(h,h) > 0, \quad \forall h \in C(\bar{x}) \setminus \{0\}, \tag{5.107}$$

其中上确界可取 $+\infty$. 事实上, 令 $(\lambda_0,\lambda) \in \Lambda^g(\bar{x})$ 满足 (5.104) 成立, 则或者 $\lambda_0 > 0$, 此时 $\lambda_0^{-1}\lambda$ 满足 (5.107), 或者 λ 是奇异的 Lagrange 乘子, 给定任何 Lagrange 乘子 $\overline{\lambda}$, 有 $\hat{\lambda} := \overline{\lambda} + t\lambda \in \Lambda(\bar{x})$ 且对充分大的 $t > 0$ 有 $\hat{\lambda}$ 满足 (5.107).

5.2.3 最优解的 Lipschitz 展式

现在讨论如下形式的参数非线性规划

$$(\mathrm{P}_u) \qquad \begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x,u), \\ \text{s. t.} \quad & g_i(x,u) = 0, \quad i = 1,\cdots,q, \\ & g_i(x,u) \leqslant 0, \quad i = q+1,\cdots,p, \end{aligned} \tag{5.108}$$

其中 $u \in U$ 是参数向量. 设对 $u = u_0$ 时的上述问题 (P_{u_0}) 与 (NLP) 问题 (5.99) 重合, U 是有限维的向量空间, $f(x,u)$ 与 $G(x,u) := (g_1(x,u),\cdots,g_p(x,u))$ 在 $\mathbb{R}^n \times U$ 上是二次连续可微的. 与 (P_u) 相联系的 Lagrange 函数是

$$L(x,\lambda,u) := f(x,u) + \sum_{i=1}^{p} \lambda_i g_i(x,u).$$

用 $\Lambda(x,u), \Lambda^g(x,u)$ 记问题 (P_u) 在点 x 处的 Lagrange 乘子集合与广义 Lagrange 乘子集合, 用 $I(x,u)$ 记在 x 处起作用的不等式约束的集合:

$$I(x,u) := \{i : g_i(x,u) = 0,\ i = q+1,\cdots,p\}.$$

对于 (P_u) 的灵敏度分析, 采用的方法是基于沿具有形式:

$$u(t) := u_0 + td + \frac{1}{2}t^2 r + o(t^2) \tag{5.109}$$

的路径 $u(\cdot) : \mathbb{R}_+ \to U$, 分析目标函数的上方与下方估计. 令 x_0 是非扰动问题 (P_{u_0}) 的最优解. 考虑在点 (x_0,u_0) 处沿方向 $d \in U$ 的 (P_u) 的如下的线性化问题

$$(\mathrm{PL}_d) \qquad \begin{aligned} \min_{h\in\mathbb{R}^n} \quad & Df(x_0,u_0)(h,d), \\ \text{s. t.} \quad & Dg_i(x_0,u_0)(h,d)=0, \quad i=1,\cdots,q, \\ & Dg_i(x_0,u_0)(h,d)\leqslant 0, \quad i\in I(x_0,u_0). \end{aligned}$$

上述问题 (PL_d) 是线性规划问题, 它的对偶问题可以写成如下的形式:

$$(\mathrm{DL}_d) \qquad \max_{\lambda\in\Lambda(x_0,u_0)} \quad D_uL(x_0,\lambda,u_0)d.$$

设 $\Lambda(x_0,u_0)$ 是非空的, 令 $\overline{\lambda}\in\Lambda(x_0,u_0)$. 则由定理 5.44 得, $\bar{h}$ 是 (PL_d) 的最优解且 $\overline{\lambda}$ 是 (DL_d) 的最优解的充分必要条件是 $\bar{h}$ 是可行的, 且

$$\overline{\lambda}_i Dg_i(x_0,u_0)(h,d)=0, \quad i\in\{1,\cdots,q\}\cup I(x_0,u_0).$$

结果, 对任何 $\overline{\lambda}\in\mathcal{S}(\mathrm{DL}_d)$,

$$\begin{aligned} \mathcal{S}(\mathrm{PL}_d)=\{h: & Dg_i(x_0,u_0)(h,d)=0, \quad i=\{1,\cdots,q\}\cup I_+(x_0,u_0,\overline{\lambda}), \\ & Dg_i(x_0,u_0)(h,d)\leqslant 0, \quad i\in I_0(x_0,u_0,\overline{\lambda})\}, \end{aligned} \tag{5.110}$$

其中

$$I_+(x_0,u_0,\overline{\lambda}):=\{i\in I(x_0,u_0):\overline{\lambda}_i>0\}, \quad I_0(x_0,u_0,\overline{\lambda}):=\{i\in I(x_0,u_0):\overline{\lambda}_i=0\}.$$

这种情形之下, 4.2 节中所讨论的方向正则性条件的抽象形式具有下述形式:

$$\left\{\begin{array}{l} D_xg_i(x_0,u_0), \quad i=1,\cdots,q \text{ 是线性无关的}, \\ \quad \exists h\in X: \quad Dg_i(x_0,u_0)(h,d)=0,\ i=1,\cdots,q, \\ \qquad\qquad\quad\ \ Dg_i(x_0,u_0)(h,d)<0,\ i\in I(x_0,u_0). \end{array}\right. \tag{5.111}$$

称上述条件为 Gollan(正则性) 条件.

给定问题 (PL_d) 的可行点 h, 记相应的起作用不等式约束集合为

$$\mathrm{I}_d(x_0,u_0,h):=\{i\in I(x_0,u_0):Dg_i(x_0,u_0)(h,d)=0\}. \tag{5.112}$$

命题 5.50 下述性质成立:

(i) Gollan 条件沿方向 d 成立当且仅当下述两个条件是成立的: (a) 向量 $D_xg_i(x_0,u_0), i=1,\cdots,q$ 是线性无关的, 且 (b) 对与 x_0 相联系的任何奇异 Lagrange 乘子 λ, 不等式 $D_uL^g(x_0,0,\lambda)d<0$ 成立.

(ii) 设 Gollan 条件沿方向 d 成立. 则 $\mathrm{val}(\mathrm{DL}_d)=\mathrm{val}(\mathrm{PL}_d)<+\infty$, 且 $\mathrm{val}(\mathrm{DL}_d)$ 是有限值的充分必要条件是集合 $\Lambda(x_0,u_0)$ 是非空的.

(iii) 若 $\Lambda(x_0,u_0)$ 是非空的, 则 Gollan 条件沿方向 d 成立当且仅当 $\mathcal{S}(\mathrm{DL}_d)$ 是非空有界的.

(iv) 若 Gollan 条件沿方向 d 成立, 则对问题 (PL_d) 的任何可行点 h, 存在 $\tilde{h}\in\mathbb{R}^n$ 满足

$$\begin{aligned}&D_xg_i(x_0,u_0)\tilde{h}=0,\quad i=1,\cdots,q,\\&D_xg_i(x_0,u_0)\tilde{h}<0,\quad i\in I_d(x_0,u_0,h).\end{aligned}\tag{5.113}$$

(v) 若 Mangasarian-Fromovitz 约束规范在点 x_0 成立, 则 Gollan 条件沿任何方向 d 均成立.

证明 结论 (i) 由命题 5.45 结论 (i) 与结论 (ii) 的等价性可立即得到.

(ii) Gollan 条件显然可推出问题 (PL_d) 是相容的. 由定理 5.44(ii) 可得 $\mathrm{val}(\mathrm{DL}_d)=\mathrm{val}(\mathrm{PL}_d)<+\infty$. 由于 $\Lambda(x_0,u_0)$ 是 (DL_d) 的可行集, 有 $\mathrm{val}(\mathrm{DL}_d)>-\infty$ 当且仅当 $\Lambda(x_0,u_0)$ 是非空的.

(iii) 若 $\Lambda(x_0,u_0)$ 是非空的且 Gollan 条件成立, 则由 (ii) 有 $\mathrm{val}(\mathrm{DL}_d)$ 是有限的, 因此由命题 5.45 可得 $\mathcal{S}(\mathrm{DL}_d)$ 是有界的. 相反地, 若 $\mathcal{S}(\mathrm{DL}_d)$ 是非空且有界的, 则 $\mathrm{val}(\mathrm{DL}_d)$ 是有限的, 且由定理 5.44(ii) 可得 $\mathrm{val}(\mathrm{DL}_d)=\mathrm{val}(\mathrm{PL}_d)$. 由命题 5.45 可得 Gollan 条件.

(iv) 令 $\bar{h}$ 是满足 (5.111) 条件的点. 显然 $\tilde{h}:=\bar{h}-h$ 满足 (5.113).

(v) 若 MF 约束规范在点 x_0 处成立, 则不存在与 x_0 相联系的奇异 Lagrange 乘子, 因此由结论 (i) 有 Gollan 条件成立. □

现在考虑的路径 $x(\cdot):\mathbb{R}_+\to\mathbb{R}^n$ 具有如下形式:

$$x(t):=x_0+th+\frac{1}{2}t^2w+o(t^2).$$

由 (5.109), f 的二阶 Taylor 展开为

$$\begin{aligned}f(x(t),u(t))=&f(x_0,u_0)+tDf(x_0,u_0)(h,d)\\&+\frac{1}{2}t^2[Df(x_0,u_0)(w,r)+D^2f(x_0,u_0)((h,d),(h,d))]+o(t^2).\end{aligned}\tag{5.114}$$

对于 $g_i(x(t),u(t))$ 也有类似的展开式. 于是有, 对所有 $t>0$ 充分小, $x(t)$ 是 $(\mathrm{P}_{u(t)})$ 的可行点, w 是下述问题的可行点:

$$\begin{array}{lll}(\mathrm{PQ}_h) & \min\limits_{w\in\mathbb{R}^n} & Df(x_0,u_0)(w,r)+D^2f(x_0,u_0)((h,d),(h,d)),\\ & \text{s. t.} & Dg_i(x_0,u_0)(w,r)+D^2g_i(x_0,u_0)((h,d),(h,d))=0,\quad i=1,\cdots,q,\\ & & Dg_i(x_0,u_0)(w,r)+D^2g_i(x_0,u_0)((h,d),(h,d))\leqslant 0,\quad i\in I_d(x_0,u_0,h),\end{array}$$

其目标函数由 $(\mathrm{P}_{u(t)})$ 的目标函数展开式的二阶项给出.

因为对任何 $h \in \mathcal{S}(\mathrm{PL}_d)$, 有

$$\mathcal{S}(\mathrm{PL}_d) = \{\lambda \in \Lambda(x_0, u_0):\ \lambda_i = 0,\ i \notin I_d(x_0, u_0, h)\},$$

问题 (PQ_h) 的对偶问题是

$$(\mathrm{DQ}_h) \qquad \max_{\lambda \in \mathcal{S}(\mathrm{DL}_d)} D_u L(x_0, \lambda, u_0) r + D^2 L(x_0, \lambda, u_0)((h, d), (h, d)).$$

令 $\mathcal{S}(\mathrm{PL}_d)$ 是问题 (PL_d) 的最优解集, 考虑与上述一对对偶问题相联系的问题:

$$(\mathcal{PQ}) \qquad \min_{h \in \mathcal{S}(\mathrm{PL}_d)} \mathrm{val}(\mathrm{PQ}_h), \tag{5.115}$$

$$(\mathcal{DQ}) \qquad \min_{h \in \mathcal{S}(\mathrm{PL}_d)} \mathrm{val}(\mathrm{DQ}_h). \tag{5.116}$$

现在给出沿路径 $u(t)$ 扰动问题最优值的某些上方估计.

定理 5.51 设 x_0 是 (P_{u_0}) 的局部最优解, 设 Gollan 条件沿方向 $d \in U$ 成立. 则

(i) 下述不等式成立

$$\limsup_{t \downarrow 0} \frac{v(u(t)) - v(u_0)}{t} \leqslant \mathrm{val}(\mathrm{PL}_d) < +\infty. \tag{5.117}$$

(ii) 若 $\mathrm{val}(\mathrm{PL}_d) > -\infty$, 则

$$\limsup_{t \downarrow 0} \frac{v(u(t)) - v(u_0) - t\ \mathrm{val}(\mathrm{PL}_d)}{\frac{1}{2}t^2} \leqslant \mathrm{val}(\mathcal{PQ}) < +\infty. \tag{5.118}$$

证明 由命题 5.50, Gollan 条件可推出 $\mathrm{val}(\mathrm{PL}_d) < +\infty$. 下证 (ii), (5.117) 中的第一不等式的证明是类似的. 设 $\mathrm{val}(\mathrm{PL}_d) > -\infty$, 因而 $\mathrm{val}(\mathrm{PL}_d)$ 是有限的. 因为 (PL_d) 是线性规划, 由定理 5.44, 它有最优解. 令 $h \in \mathcal{S}(\mathrm{PL}_d)$, $\bar{h}$ 与 $\tilde{h}$ 是满足 Gollan 条件 (5.111) 与命题 5.50 的条件 (5.113) 的点. 以下验证 (PQ_h) 是可行的. 因为等式约束的导数是线性无关的, 存在 w_0 满足 (PQ_h) 的等式约束, 则对充分大的 $t > 0$, 由 (5.113) 显然可以得到 $w_0 + t\tilde{h}$ 是 (PQ_h) 的可行点. 问题 (PQ_h) 的可行性可推出 $\mathrm{val}(\mathrm{PQ}_h) = \mathrm{val}(\mathrm{DQ}_h)$ 对所有的 $h \in \mathcal{S}(\mathrm{PL}_d)$ 均成立, 因此 $\mathrm{val}(\mathcal{PQ}) = \mathrm{val}(\mathcal{DQ})$.

令 w 是 (PQ_h) 的可行点. 给定 $\varepsilon > 0$, 置 $w_\varepsilon := w + \varepsilon\bar{h}$, 有 w_ε 是 (PQ_h) 的可行点且相应的不等式约束在 w_ε 处是严格负的. 类似于命题 5.47 的证明, 存在路径 $x(t) := x_0 + th + \frac{1}{2}t^2 w_\varepsilon + o(t^2)$ 满足对充分小的 $t > 0$, $x(t)$ 是 $(\mathrm{P}_{u(t)})$ 的可行点. 从而有 $\mathrm{val}(\mathrm{P}_{u(t)}) \leqslant f(x(t), u(t))$. 因为

$$\begin{aligned} f(x(t), u(t)) = {} & f(x_0, u_0) + tDf(x_0, u_0)(h, d) + \frac{1}{2}t^2[Df(x_0, u_0)(w_\varepsilon, r) \\ & + D^2 f(x_0, u_0)((h, d), (h, d))] + o(t^2), \end{aligned}$$

用不等式 $v(u(t)) \leqslant f(x(t), u(t))$ 与等式 $f(x_0, u_0) = \text{val}(\text{P}_{u_0})$, $Df(x_0, u_0)(h, d) = \text{val}(\text{PL}_d)$, 得到

$$\limsup_{t \downarrow 0} \frac{v(u(t)) - v(u_0) - t\ \text{val}(\text{PL}_d)}{\frac{1}{2}t^2} \leqslant Df(x_0, u_0)(w_\varepsilon, r) + D^2 f(x_0, u_0)((h, d), (h, d)).$$

取 $\varepsilon \downarrow 0$ 时的极限, 用到 w 是 (PQ_h) 的任意可行点这一事实, 得到 (5.118) 的左端小于或等于 $\text{val}(\text{PQ}_h)$. 对 $h \in \mathcal{S}(\text{PL}_d)$ 取下确界, 可得到不等式 (5.118). □

注 5.52 由上述定理的第一估计式可得, 若 Gollan 条件沿方向 d 成立, 且 Lagrange 乘子集合 $\Lambda(x_0, u_0)$ 是空集, 则

$$\lim_{t \downarrow 0} \frac{v(u(t)) - v(u_0)}{t} = -\infty. \tag{5.119}$$

现在考虑 (沿方向 d 的) 二阶充分最优性条件的下述强形式

$$\sup_{\lambda \in \mathcal{S}(\text{DL}_d)} D^2_{xx} L(x_0, \lambda, u_0)(h, h) > 0, \quad \forall h \in C(x_0) \setminus \{0\}. \tag{5.120}$$

因为 $\mathcal{S}(\text{DL}_d)$ 是 $\Lambda(x_0, u_0)$ 的子集, 这一条件显然要比 "标准" 条件 (5.104) 或 (5.107) 强. 现在叙述这一节的主要结果.

定理 5.53 设

(i) 非扰动问题 (P_{u_0}) 有唯一的最优解 x_0.

(ii) Gollan 条件沿方向 d 成立.

(iii) Lagrange 乘子集合 $\Lambda(x_0, u_0)$ 是非空的.

(iv) 强二阶充分性条件(5.120) 成立.

(v) 对所有 $t > 0$ 充分小, $(\text{P}_{u(t)})$ 的可行集是非空的且一致有界的.

则对 $(\text{P}_{u(t)})$ 的任何 $o(t^2)$ 最优解 $\bar{x}(t)$, 其中 $t \geqslant 0$, 下述结论成立:

(a) $\bar{x}(t)$ 在 x_0 处是 Lipschitz 稳定的, 即 $\|\bar{x} - x_0\| = O(t)$.

(b) 问题 $(\mathcal{PQ})$ 具有有限的最优值, 且最优值函数有下述的展式成立:

$$v(u(t)) = v(u_0) + t\ \text{val}(\text{PL}_d) + \frac{1}{2}t^2\, \text{val}(\mathcal{PQ}) + o(t^2), \tag{5.121}$$

(c) $(\bar{x}(t) - x_0)/t$ 的任何极限点均是 $(\mathcal{PQ})$ 的最优解.

(d) $(\mathcal{PQ})$ 的任一最优解 h 均联系着具有 $x(t) = x_0 + th + o(t)$ 形式的路径满足 $x(t)$ 是 $(\text{P}_{u(t)})$ 的 $o(t^2)$ 最优解.

(e) 与 $\bar{x}(t)$ 相联系的 Lagrange 乘子的任何极限点属于 $\mathcal{S}(\text{DL}_d)$.

证明 (a) 假设 (v) 可推出 $\bar{x}(t)$ 是有界的, 因而当 $t \downarrow 0$ 时有极限点 $\bar{x}$, 则由定理 5.51,

$$f(\bar{x}, u_0) \leqslant \limsup_{t \downarrow 0} f(\bar{x}(t), u(t)) \leqslant v(u_0),$$

因为映射 $G(x,u)$ 是连续的, 所以 $\bar{x}$ 是 (P_{u_0}) 的可行点. 得到 $\bar{x}$ 是 (P_{u_0}) 的最优解, 因此 $\bar{x}=x_0$. 证得当 $t\downarrow 0$ 时 $\bar{x}(t)\to x_0$.

现在证 $\bar{x}(t)$ 在 x_0 处是 Lipschitz 稳定的. 假设不真. 则存在序列 $t_k\downarrow 0$ 满足, 记 $u_k:=u(t_k)$, $x_k:=\bar{x}(t_k)$ 与 $\tau_k:=\|x_k-x_0\|$, 有 x_k 是 (P_{u_k}) 的 $o(t_k^2)$ 最优解且 $\tau_k/t_k\to+\infty$. 如有必要, 可以抽取一子列, 不妨设 $h_k:=(x_k-x_0)/\tau_k$ 收敛到某一向量 $\bar{h}$. 注意到 $\|h_k\|=1$, 有 $\|\bar{h}\|=1$. 由 $G(x_k,u_k)$ 的一阶展式, 有

$$D_xg_i(x_0,u_0)\bar{h}=0,\quad i=1,\cdots,q\quad 且\ D_xg_i(x_0,u_0)\bar{h}\leqslant 0,\quad i\in I(x_0,u_0).$$

结合 $f(x_k,u_k)$ 的一阶展开式与一阶上方近似 (5.117), 可推出 $\bar{h}$ 是非零的临界方向. 由 (5.120), 存在 $\lambda\in\mathcal{S}(\mathrm{DL}_d)$ 满足 $D_{xx}^2L(x_0,\lambda,u_0)(\bar{h},\bar{h})$ 是正值的. 因为 $D_xL(x_0,\lambda,u_0)=0$ 且

$$D_uL(x_0,\lambda,u_0)d=\mathrm{val}(\mathrm{DL}_d)=\mathrm{val}(\mathrm{PL}_d),\tag{5.122}$$

以及 $\lambda_i\geqslant 0$ 对所有的 $i=q+1,\cdots,p$ 是成立的, 所以得到

$$\begin{aligned}f(x_k,u_k)&\geqslant L(x_k,\lambda,u_k)\\&=L(x_0,\lambda,u_0)+t_kD_uL(x_0,\lambda,u_0)d+\frac{1}{2}\tau_k^2D_{xx}^2L(x_0,\lambda,u_0)(\bar{h},\bar{h})+o(\tau_k^2)\\&=f(x_0,u_0)+t_k\ \mathrm{val}(\mathrm{PL}_d)+\frac{1}{2}\tau_k^2D_{xx}^2L(x_0,\lambda,u_0)(\bar{h},\bar{h})+o(\tau_k^2).\end{aligned}$$

这导致

$$\lim_{k\to\infty}\frac{v(u_k)-v(u_0)-t_k\ \mathrm{val}(\mathrm{PL}_d)}{t_k^2}=+\infty.$$

然而这与定理 5.51 的第二估计式是相矛盾的. 证得 (a).

(b) 令 $t_k\downarrow 0$ 是序列, 沿此序列

$$\frac{v(t)-v(u_0)-t\ \mathrm{val}(\mathrm{PL}_d)}{t^2}$$

的下极限被取到. 置 $x_k:=\bar{x}(t_k)$, $u_k:=u(t_k)$. 如有必要, 可抽取一子列, 不妨设 $(x_k-x_0)/t_k$ 收敛到某一向量 $\bar{h}$. 由 $G(x_k,u_k)$ 的一阶展开式, 可得 $\bar{h}$ 是 (PL_d) 的可行点. 将之与 $f(x_k,u_k)$ 的一阶展式与定理 5.51 结合, 得到 $\bar{h}\in\mathcal{S}(\mathrm{PL}_d)$. 固定 $\lambda\in\mathcal{S}(\mathrm{DL}_d)$. 用 $L(x_0,\lambda,u_0)=f(x_0,u_0)$ 与 (5.122), 得到

$$\begin{aligned}f(x_k,u_k)\geqslant&\,L(x_k,\lambda,u_k)\\=&\,f(x_0,u_0)+t_k\ \mathrm{val}(\mathrm{PL}_d)+\frac{1}{2}t_k^2[D_uL(x_0,\lambda,u_0)r\\&+D^2L(x_0,\lambda,u_0)((\bar{h},d)(\bar{h},d))]+o_\lambda(t_k^2).\end{aligned}$$

用 $o_\lambda(\cdot)$ 记一项, 满足对每一 $\lambda \in \mathcal{S}(\text{DL}_d)$, 有 $\|o_\lambda(t_k^2)\|/t_k^2 \to 0$. 另外, 因为集合 $\mathcal{S}(\text{DL}_d)$ 是有界的, 有 $\|o_\lambda(t_k^2)\|/t_k^2 \to 0$ 对 $\lambda \in \mathcal{S}(\text{DL}_d)$ 是一致的. 因为 $\bar{h}$ 是 (PL_d) 的可行点, 有

$$D_u L(x_0, \lambda, u_0)r + D^2 L(x_0, \lambda, u_0)((\bar{h}, d)(\bar{h}, d)) \geqslant \text{val}(\mathcal{DQ}). \tag{5.123}$$

同时, 有 $\text{val}(\mathcal{DQ}) = \text{val}(\mathcal{PQ})$, 而由定理 5.51(ii) 有 (5.123) 的左端小于等于 $\text{val}(\mathcal{PQ})$. 结论 (b) 得证.

(c) 因为 (5.121) 成立, 上述推证可推出 (5.123) 的左端项等于 $\text{val}(\mathcal{PQ})$, 因此有 $\bar{h}$ 是 $(\mathcal{PQ})$ 的最优解.

(d) 设 h 是 $(\mathcal{PQ})$ 的最优解, 则线性规划 (PQ_h) 与其对偶是相容的. 结果, 存在最优解 $w \in \mathcal{S}(\text{PQ}_h)$ 与 $\lambda \in \mathcal{S}(\text{DQ}_h)$. 令 $\tilde{h}$ 是满足条件 (5.113) 的一点 (它的存在性由 Gollan 条件保证), 考虑路径 $\hat{x}_\varepsilon(t) := x_0 + th + \frac{1}{2}t^2(w + \varepsilon\tilde{h})$, 其中 ε 是小的正的参数. 用二阶 Taylor 展式得 $g_i(\hat{x}_\varepsilon(t), u(t)) = o(t^2), i = 1, \cdots, q$. 则由隐函数定理, 存在一路径 $x_\varepsilon(t) = \hat{x}_\varepsilon(t) + o(t^2)$ 满足 $g_i(x_\varepsilon(t), u(t)) = o(t^2)$, 对 $i = 1, \cdots, q$ 成立. 还有

$$g_i(x_\varepsilon(t), u(t)) \leqslant \frac{1}{2}\varepsilon t^2 D_x g_i(x_0, u_0)\tilde{h} + o(t^2), \quad i \in I_d(x_0, u_0, h).$$

因此, 对所有充分小的 $t > 0$, $x_\varepsilon(t)$ 是 $(\text{P}_{u(t)})$ 的可行点. 由 $f(x_\varepsilon(t), u(t))$ 的二阶展式, 由 (5.121) 得

$$\begin{aligned}
f(x_\varepsilon(t), u(t)) &= f(x_0, u_0) + tDf(x_0, u_0)(h, d) \\
&\quad + \frac{1}{2}t^2[Df(x_0, u_0)(w + \varepsilon\tilde{h}, r) + D^2 f(x_0, u_0)((h, d), (h, d))] + o(t^2) \\
&= f(x_0, u_0) + t\ \text{val}(\text{PL}_d) + \frac{1}{2}\varepsilon t^2 D_x f(x_0, u_0)\tilde{h} \\
&\quad + \frac{1}{2}t^2[D_u L(x_0, \lambda, u_0)r + D^2 L(x_0, \lambda, u)((h, d), (h, d))] + o(t^2) \\
&= v(u(t)) + \frac{1}{2}\varepsilon t^2 D_x f(x_0, u_0)\tilde{h} + o(t^2).
\end{aligned}$$

令 $t_\varepsilon > 0$ 满足

$$\left\|x_0 + th + \frac{1}{2}t^2 w - x_\varepsilon(t)\right\| \leqslant 2\varepsilon^{1/2}(\|\tilde{h}\| + 1)t^2, \quad \forall t \leqslant t_\varepsilon.$$

这对 $\varepsilon \in (0, 1)$ 是有定义的, 因为

$$x_\varepsilon(t) = x_0 + th + \frac{1}{2}t^2[w + \varepsilon\tilde{h}] + o(t^2).$$

考虑由下述迭代关系定义的序列 τ_n:

$$\tau_1 = t_1, \quad \tau_{n+1} = \min\left(\frac{1}{2}\tau_n,\ t_1/(n+1)\right),$$

置 $x(t) := x_{1/n}(t),\ t \in [\tau_{n+1}, \tau_n)$, 则

$$\left\| x_0 + th + \frac{1}{2}t^2 w - x(t) \right\| \leqslant 2n^{-1/2}(\|\tilde{h}\| + 1)t^2$$

对所有 $t < \tau_n$ 成立, 因此 $x(t) = x_0 + th + \frac{1}{2}t^2 w + o(t^2)$, 由 $f(x_\varepsilon(t), u(t))$ 的二阶展式, 即得所需要的结果.

(e) 由命题 (c), $(\bar{x}(t) - x_0))/t$ 的任何极限点 $\bar{h}$ 属于 $\mathcal{S}(\mathrm{PL}_d)$. 显然, 若 $i \in I(x_0, u_0)$ 满足 $Dg_i(x_0, u_0)(\bar{h}, d) < 0$, 则对充分小的 $t > 0$, $g_i(\bar{x}(t), u(t)) < 0$, 因此, 若 $\lambda_t \in \Lambda(\bar{x}(t), u(t))$, 则 $(\lambda_t)_i = 0$. 所以, λ_t 的任何极限点 $\overline{\lambda}$ 满足

$$\bar{\lambda}_i Dg_i(x_0, u_0)(\bar{h}, d) = 0, \quad \forall i \in I(x_0, u_0). \tag{5.124}$$

在 $\Lambda(x_t, u(t))$ 的定义中取极限, 容易验证 $\overline{\lambda} \in \Lambda(x_0, u_0)$. 由于 $\mathcal{S}(\mathrm{DL}_d)$ 恰是满足 (5.124) 的 $\Lambda(x_0, u_0)$ 的子集, 结论 (e) 证得. □

注 5.54 上述定理的假设 (v) 只是用于保证 $\bar{x}(t)$ 是有界的, 因此, 当 $t \downarrow 0$ 时 $\bar{x}(t) \to x_0$. 可以用下确界紧致性条件来代替这一假设, 或可用确保当 $t \downarrow 0$ 时 $\bar{x}(t) \to x_0$ 的某一事先的条件来代替这一假设.

注 5.55 令 $\bar{x}(u)$ 是 (P_u) 的 $o(\|u - u_0\|^2)$ 最优解. 设定理 5.53 的假设 (ii)~(iv) 及确保当 $t \downarrow 0$ 时 $\bar{x}(u_0 + td)$ 趋于 x_0 的假设成立, 则 $\bar{x}(u)$ 在 u_0 处沿方向 d 是方向可微的, 其方向导数$\bar{x}'(u_0, d)$ 由下述问题的最优解 $\bar{h}$ 给出:

$$\min_{h \in \mathcal{S}(\mathrm{PL}_d)} \Big\{ \max_{\lambda \in \mathcal{S}(\mathrm{DL}_d)} D^2 L(x_0, \lambda, u_0)((h, d), (h, d)) \Big\}, \tag{5.125}$$

若这一最优解 $\bar{h}$ 是唯一的. 因为在假设 (ii) 与 (iii) 之下, $\mathcal{S}(\mathrm{DL}_d)$ 是非空的凸的紧致的多面集, 由于公式 (5.110), 由如下的二阶条件:

$$\begin{aligned} &\text{对任何 } \lambda \in \mathcal{S}(\mathrm{DL}_d), \text{ 对满足 } D_x g_i(x_0, u_0)h = 0, i \in \{1, \cdots, q\} \cup I_+(x_0, u_0, \lambda) \\ &\qquad \text{的 } h \neq 0, \ \text{有} \ \ D^2_{xx} L(x_0, \lambda, u_0)(h, h) > 0, \end{aligned} \tag{5.126}$$

这样的最优解 $\bar{h}$ 的唯一性是确保的. 事实上, 对任何 $\lambda \in \Lambda(x_0, u_0)$, 由集合 $\mathcal{S}(\mathrm{PL}_d)$ 生成的线性空间包含在由 (5.126) 的线性约束所定义的线性空间中, 在此线性空间上, 由 $D^2_{xx} L(x_0, \lambda, u_0)$ 的正定性可推出相应的二次函数在此空间上是严格凸的.

5.2.4 最优解的 Hölder 展式

这一节讨论如下两种情形的由 (5.108) 定义的参数化问题 (P_u) 的扰动分析. 第一种情况是 Lagrange 乘子的集合 $\Lambda(x_0, u_0)$ 非空但强二阶充分条件 (5.120) 不成立

的情况; 第二种情况是 $\Lambda(x_0,u_0)$ 是空集的情况. 这一节用如下形式的路径 $u(\cdot):\mathbb{R}_+\to U$ 与 $x(\cdot):\mathbb{R}_+\to\mathbb{R}^n$,

$$u(t):=u_0+td+o(t), \tag{5.127}$$

$$x(t):=x_0+t^{1/2}h+tw+o(t), \tag{5.128}$$

其中 h 是临界方向, 即 $h\in C(x_0)$.

先设 $\Lambda(x_0,u_0)$ 是非空的. 定义

$$I(x_0,u_0,h):=\{i\in I(x_0,u_0):D_xg_i(x_0,u_0)h=0\}.$$

由二阶 Taylor 展式, 有

$$\begin{aligned}f(x(t),u(t))=&f(x_0,u_0)+t^{1/2}D_xf(x_0,u_0)h\\&+\frac{1}{2}t[2Df(x_0,u_0)(w,d)+D^2_{xx}f(x_0,u_0)(h,h)]+o(t),\end{aligned} \tag{5.129}$$

类似地, 函数 $g_i(x(t),u(t))$ 也有上述的二阶展式. 可得到, 若对 $t>0$ 充分小, $x(t)$ 是 $(\mathrm{P}_{u(t)})$ 的可行点, 则 w 是下述问题的可行点

$$\begin{array}{llll}(\mathrm{PQ}^2_h) & \min\limits_{w\in\mathbb{R}^n} & 2Df(x_0,u_0)(w,d)+D^2_{xx}f(x_0,u_0)(h,h), & \\ & \text{s. t.} & 2Dg_i(x_0,u_0)(w,d)+D^2_{xx}g_i(x_0,u_0)(h,h)=0, & i=1,\cdots,q,\\ & & 2Dg_i(x_0,u_0)(w,d)+D^2_{xx}g_i(x_0,u_0)(h,h)\leqslant 0, & i\in I(x_0,u_0,h).\end{array}$$

注意到, 因为 h 是临界方向, 设 $\Lambda(x_0,u_0)$ 是非空的, 有 $D_xf(x_0,u_0)h=0$. 所以, 上述问题 (PQ^2_h) 的目标函数由展式 (5.129) 的一阶项给出.

还有, 对每一 $\lambda\in\Lambda(x_0,u_0)$, 对所有的 $i\in\{q+1,\cdots,p\}\setminus I(x_0,u_0,h)$, $\lambda_i=0$. 问题 (PQ^2_h) 的对偶问题可以表示为如下的形式:

$$(\mathrm{DQ}^2_h)\qquad \max_{\lambda\in\Lambda(x_0,u_0)} 2D_uL(x_0,\lambda,u_0)d+D^2_{xx}L(x_0,\lambda,u_0)(h,h).$$

对所有的临界方向极小化上述问题的最优值, 得到下述两个问题:

$$(\mathcal{PQ}^2)\qquad \min_{h\in C(x_0)}\ \mathrm{val}(\mathrm{PQ}^2_h), \tag{5.130}$$

$$(\mathcal{DQ}^2)\qquad \min_{h\in C(x_0)}\ \mathrm{val}(\mathrm{DQ}^2_h). \tag{5.131}$$

因为 $\Lambda(x_0,u_0)\neq 0$, 问题 (DQ^2_h) 是相容的. 由定理 5.44(ii) 可得 $\mathrm{val}(\mathrm{DQ}^2_h)=\mathrm{val}(\mathrm{PQ}^2_h)$. 对临界方向 $h=0$, 问题 (DQ^2_h), 相差倍数 2, 与 (DL_d) 是重合的. 因为在 Gollan 条件下, (DL_d) 的最优值等于 (PL_d) 的最优值, 可得到

$$\mathrm{val}(\mathcal{DQ}^2)=\mathrm{val}(\mathcal{PQ}^2)\leqslant 2\,\mathrm{val}(\mathrm{PL}_d). \tag{5.132}$$

现在叙述最优值函数的一阶上方近似，它强于上方近似 (5.117).

定理 5.56　设 x_0 是 (P_{u_0}) 的局部最优解，$\Lambda(x_0,u_0)$ 是非空的且沿方向 $d\in U$ 的 Gollan 条件成立，则

$$\limsup_{t\downarrow 0}\frac{v(u(t))-v(u_0)}{t}\leqslant\frac{1}{2}\operatorname{val}(\mathcal{P}Q^2)\leqslant\operatorname{val}(\mathrm{PL}_d)<+\infty. \tag{5.133}$$

证明　(5.133) 的后一不等式可根据定理 5.51, 由 Gollan 条件与 (5.132) 之前得到的那个不等式得到. 因此, 只需证明 (5.133) 的第一个不等式.

令 h 是临界方向且满足 (PQ_h^2) 是相容的，设 w 是 (PQ_h^2) 的可行点，$\hat{h}$ 是满足 Gollan 条件(5.111) 的一点. 令 $\gamma\in(0,1)$. 用 γ^2 乘以 $(\mathrm{PQ}_{\gamma h}^2)$ 的约束加上 (5.111) 中线性关系的 $(1-\gamma^2)$ 倍，得到 $w_\gamma:=\gamma^2w+(1-\gamma^2)\hat{h}$ 是 $(\mathrm{PQ}_{\gamma h}^2)$ 的可行点，且严格地满足不等式约束. 令 $\hat{x}(t):=x_0+\gamma t^{1/2}h+tw_\gamma$, 由于

$$\begin{aligned}G(\hat{x}(t),u(t))=&G(x_0,u_0)+\gamma t^{1/2}D_xG(x_0,u_0)h\\&+\frac{1}{2}t[2DG(x_0,u_0)(w_\gamma,d)+\gamma^2D^2G(x_0,u_0)(h,h)]+o(t)\\=&G(x_0,u_0)+\gamma t^{1/2}D_xG(x_0,u_0)h\\&+\frac{1}{2}t\gamma^2[2DG(x_0,u_0)(w,d)+D^2G(x_0,u_0)(h,h)]\\&+(1-\gamma^2)tDG(x_0,u_0)(\hat{h},d)+o(t),\end{aligned}$$

有

$$\begin{aligned}&g_i(\hat{x}(t),u(t))=o(t),\quad i=1,\cdots,q,\\&g_i(\hat{x}(t),u(t))\leqslant(1-\gamma^2)tDg_i(x_0,u_0)(\hat{h},d)+o(t),\quad i\in I(x_0,u_0,h).\end{aligned}$$

则由隐函数定理得, 存在路径 $x(t)=\hat{x}(t)+o(t)$ 满足对 $i=1,\cdots,q$, $g_i(x(t),u(t))=0$, 且对于 $i\in I(x_0,u_0,h)$, 得到

$$g_i(x(t),u(t))\leqslant(1-\gamma^2)tDg_i(x_0,u_0)(\hat{h},d)+o(t)<0$$

对 $t>0$ 充分小成立. 得到

$$\begin{aligned}\limsup_{t\downarrow 0}\frac{v(u(t))-v(u_0)}{t}&\leqslant\lim_{t\downarrow 0}\frac{f(x(t),u(t))-v(u_0)}{t}\\&=Df(x_0,u_0)(w_\gamma,d)+\frac{1}{2}D^2f(x_0,u_0)(h,h).\end{aligned}$$

取 $\gamma\uparrow 1$ 时的极限, 得到

$$\limsup_{t\downarrow 0}\frac{v(u(t))-v(u_0)}{t}\leqslant Df(x_0,u_0)(w,d)+\frac{1}{2}D^2f(x_0,u_0)(h,h).$$

因为 w 是 (PQ_h^2) 的任意的可行点, 得到

$$\limsup_{t\downarrow 0}\frac{v(u(t))-v(u_0)}{t}\leqslant\frac{1}{2}\ \mathrm{val}(\mathrm{PQ}_h^2),$$

因此 (5.133) 的第一不等式成立. □

下述定理给出 Lagrange 乘子集合是非空情况下基本的灵敏性结果.

定理 5.57 设

(i) 非扰动问题 (P_{u_0}) 有唯一的最优解 x_0.

(ii) 沿方向 d 的 Gollan 条件成立.

(iii) Lagrange 乘子集合 $\Lambda(x_0,u_0)$ 是非空的.

(iv) 二阶充分性条件 (5.107) 成立.

(v) 对所有的充分小的 $t>0$, $(\mathrm{P}_{u(t)})$ 的可行集是非空的且是一致有界的.

则对 $(\mathrm{P}_{u(t)})$ 的任何 $o(t)$ 最优解 $\bar{x}(t)$, 其中 $t\geqslant 0$, 下述结论成立:

(a) $\bar{x}(t)$ 在 x_0 处是 Hölder 稳定的, 即 $\|\bar{x}(t)-x_0\|=O(t^{1/2})$.

(b) 最优值函数具有下述的展开式

$$v(u(t))=v(u_0)+\frac{1}{2}t\ \mathrm{val}(\mathcal{P}Q^2)+o(t). \tag{5.134}$$

(c) $t^{-1/2}(\bar{x}(t)-x_0)$ 的任何极限点是 $(\mathcal{P}Q^2)$ 的最优解.

(d) 与 $(\mathcal{P}Q^2)$ 的任何最优解 h 相联系着形式为 $x(t)=x_0+t^{1/2}h+o(t^{1/2}),t>0$ 的路径, 满足 $x(t)$ 是 $(\mathrm{P}_{u(t)})$ 的 $o(t)$ 最优解.

证明 (a) 设 $\bar{x}(t)$ 是 $(\mathrm{P}_{u(t)})$[①]的 $o(t)$ 最优解. 由类似于定理 5.53(a) 的推证得到, 当 $t\downarrow 0$ 时 $\bar{x}(t)\to x_0$.

以下证明 $\bar{x}(t)$ 在 x_0 处是 Hölder 稳定的. 在 Hölder 稳定性的下述证明中, 我们不用假设 (iii), 用条件 (5.104) 代替条件 (5.107). 在最优解处的广义 Lagrange 乘子集合总是非空的, 且若 $\Lambda(x_0,u_0)$ 非空, 则二阶条件 (5.104) 与 (5.107) 是等价的.

假设 Hölder 稳定性不成立, 则存在序列 $t_k\downarrow 0$ 满足, 若置 $u_k:=u(t_k),x_k:=\bar{x}(t_k),\tau_k:=\|x_k-x_0\|$, 有 $\tau_k^2/t_k\to+\infty$. 如有必要, 可抽取一子列, 可设 $h_k:=(x_k-x_0)/\tau_k$ 收敛到向量 $\bar{h}$, 它具有单位范数. 由 $G(x_k,u_k)$ 与 $f(x_k,u_k)$ 的一阶展开式及一阶上方近似 (5.117), 有 $\bar{h}$ 是临界方向. 由二阶充分性条件 (5.104), 存在 $(\lambda_0,\lambda)\in\Lambda^g(x_0,u_0)$ 满足 $D_{xx}^2L^g(x_0,\lambda_0,\lambda,u_0)(\bar{h},\bar{h})$ 是正的, 则有

$$\begin{aligned}\lambda_0 f(x_k,u_k)&\geqslant L^g(x_k,\lambda_0,\lambda,u_k)\\&=\lambda_0 f(x_0,u_0)+\frac{1}{2}\tau_k^2D_{xx}^2L^g(x_0,\lambda_0,\lambda,u_0)(\bar{h},\bar{h})+o(\tau_k^2).\end{aligned}$$

若 $\lambda_0=0$, 这导致矛盾; 否则, 有

$$\lim_{k\to\infty}\frac{v(u_k)-v(u_0)}{t_k}=+\infty,$$

① 原著中为 $S(\mathrm{P}_{u(t)})$.

它与定理 5.51 是矛盾的.

(b) 由定理 5.56, 只需证 (5.134) 的左端大于或等于其右端, 且 $\mathrm{val}(\mathcal{PQ}^2) > -\infty$. 设 $t_k \downarrow 0$ 是序列, 沿此序列的极限, $[v(u(t)) - v(u_0)]/t$ 的下极限被取到. 置 $x_k := \bar{x}(t_k), u_k := u(t_k)$. 因为 $\|x_k - x_0\| = O(t_k^{1/2})$, 如有必要可抽取一子列, 不妨设 $t_k^{-1/2}(x_k - x_0)$ 收敛到向量 $\bar{h}$, 有 $\bar{h} \in C(x_0)$.

令 $\lambda \in \Lambda(x_0, u_0)$. 因为 $\bar{h} \in C(x_0)$, 有

$$\begin{aligned}f(x_k, u_k) &\geqslant L(x_k, \lambda, u_k)\\ &= f(x_0, u_0) + \frac{1}{2}t_k[2D_u L(x_0, \lambda, u_0)d + \frac{1}{2}D_{xx}^2 L(x_0, \lambda, u_0)(\bar{h}, \bar{h})] + o(t_k),\end{aligned}$$

因此

$$\liminf_{k\to\infty} \frac{f(x_k, u_k) - f(x_0, u_0)}{t_k} \geqslant D_u L(x_0, \lambda, u_0)d + \frac{1}{2}D_{xx}^2 L(x_0, \lambda, u_0)(\bar{h}, \bar{h}).$$

对 λ 于 $\Lambda(x_0, u_0)$ 上取极大值, 得到上式左端大于或等于 $\mathrm{val}(\mathcal{DQ}^2)$. 因为 $\mathrm{val}(\mathcal{DQ}^2) = \mathrm{val}(\mathcal{PQ}^2)$, 即证得所需要的不等式.

(c) 因为 (5.134) 成立, 上述推证也可推出

$$2D_u L(x_0, \overline{\lambda}, u_0)d + D_{xx}^2 L(x_0, \overline{\lambda}, u_0)(\bar{h}, \bar{h}) = \mathrm{val}(\mathcal{DQ}^2),$$

因此 $\bar{h}$ 是 $(\mathcal{PQ}^2)$ 的最优解.

(d) 给定 $h \in \mathcal{S}(\mathcal{PQ}^2)$, 则线性规划 (PQ_h^2) 以及它的对偶 (DQ_h^2) 是相容的, 因此存在最优解 $w \in \mathcal{S}(\mathrm{PQ}_h^2)$. 给定 $\gamma \in (0,1)$, 置 $w_\gamma := \gamma^2 w + (1-\gamma^2)\hat{h}$. 在定理 5.56 的证明证得存在形式为 $x_\gamma(t) := x_0 + \gamma t^{1/2}h + tw_\gamma + o(t)$ 的路径, 满足对所有充分小的 $t > 0$, $x_\gamma(t)$ 是 $(\mathrm{P}_{u(t)})$ 的可行点. 由 $f(x_\gamma(t), u(t))$ 的 Taylor 展式, 用到 (5.121) 得到

$$\begin{aligned}f(x_\gamma(t), u(t)) &= f(x_0, u_0) + \frac{1}{2}t[2Df(x_0, u_0)(w_\gamma, d) + \gamma^2 D^2 f(x_0, u_0)(h, h)] + o(t)\\ &= f(x_0, u_0) + \gamma^2 t\ \mathrm{val}(\mathrm{PQ}_h^2) + (1-\gamma^2)tDf(x_0, u_0)(\hat{h}, d) + o(t)\\ &= \mathrm{val}\ (\mathrm{P}_{u(t)}) + (1-\gamma^2)t[D_x f(x_0, u_0)\hat{h} - \mathrm{val}(\mathrm{PQ}_h^2)] + o(t).\end{aligned}$$

令 $t_\gamma > 0$ 满足

$$\|x_0 + t^{1/2}h - x_\gamma(t)\| \leqslant 2(1-\gamma)(\|h\| + 1)t^{1/2}, \quad \forall t \leqslant t_\gamma$$

(译者注：上式中原著为 t_ε), $x_\gamma(t) = x_0 + \gamma t^{1/2}h + o(t)$, 这是有定义的. 考虑序列 $\gamma_n := 1 - 1/n$ 与 τ_n, 其中后者由迭代关系定义：$\tau_1 = t_1, \tau_{n+1} = \min\left(\frac{1}{2}\tau_n, t_{\gamma_n}\right)$, 置 $x(t) := x_{\gamma_n}(t)$, $t \in [\tau_{n+1}, \tau_n)$, 则对所有的 $t < \tau_n$,

$$\|x_0 + t^{1/2}h - x(t)\| \leqslant 2n^{-1}(\|h\| + 1)t,$$

从而有 $x(t)=x_0+t^{\frac{1}{2}}h+o(t^{\frac{1}{2}})$. 将 $f(x(t),u(t))$ 展开, 即得到结果. □

注意到, 展开式 (5.134) 意味着在上述定理的假设下, 最优值函数在点 u_0 处沿方向 d 是 Hadamard 意义下方向可微的, 且

$$v'(u_0,d)=\frac{1}{2}\operatorname{val}(\mathcal{P}\mathcal{Q}^2). \tag{5.135}$$

最后讨论 Lagrange 乘子集合 $\Lambda(x_0,u_0)$ 是空集, Gollan 条件沿方向 d 成立的情况. 在这种情况下, 线性化问题 (PL_d) 是相容的, 因为 $\Lambda(x_0,u_0)$ 是空集, 有 $\operatorname{val}(\mathrm{PL}_d)=\operatorname{val}(\mathrm{DL}_d)=-\infty$. 于是有 (见注 5.52)

$$\lim_{t\downarrow 0}\frac{v(u(t))-v(u_0)}{t}=-\infty.$$

注意到, 因为 $\operatorname{val}(\mathrm{PL}_d)=-\infty$, 存在临界方向 $h\in C(x_0)$ 满足 $D_xf(x_0,u_0)h<0$. 对形式 (5.128) 的路径, $f(x(t),u(t))$ 展式的第一个非负项是 $t^{1/2}D_xf(x_0,u_0)h$. 所以, 考虑下述的辅助问题:

$$\begin{array}{lll}(\mathrm{PQ}_h^3) & \min\limits_{w\in\mathbb{R}^n} & D_xf(x_0,u_0)h,\\ & \text{s. t.} & 2Dg_i(x_0,u_0)(w,d)+D_{xx}^2g_i(x_0,u_0)(h,h)=0,\quad i=1,\cdots,q,\\ & & 2Dg_i(x_0,u_0)(w,d)+D_{xx}^2g_i(x_0,u_0)(h,h)\leqslant 0,\quad i\in I(x_0,u_0,h).\end{array}$$

这一问题类似于问题 (PQ_h^2). 然而, 主要的区别是变量 w 不进入 (PQ_h^3) 的目标函数, 因此这一极小化问题实际上是可行性问题.

由于假设 $\Lambda(x_0,u_0)$ 是空集, 相应的广义 Lagrange 乘子集合 $\Lambda^g(x_0,u_0)$ 只含有奇异的 Lagrange 乘子. 用 $\Lambda^s(x_0,u_0)$ 记满足下述最优性条件的奇异 Lagrange 乘子$\lambda\neq 0$ 的集合:

$$D_xL^s(x_0,\lambda,u_0)=0,\quad \lambda_i\geqslant 0,\quad \lambda_ig_i(x_0,u_0)=0,\quad i=q+1,\cdots,p, \tag{5.136}$$

其中 $L^s(x,\lambda,u):=\sum\limits_{i=1}^{p}\lambda_ig_i(x,u)$. 注意到 $\Lambda^s(x_0,u_0)\cup\{0\}$ 是凸的多面体锥.

对临界方向 $h\in C(x_0)$ 及每一 $\lambda\in\Lambda^s(x_0,u_0)$, 对所有的 $i\in\{q+1,\cdots,p\}\setminus I(x_0,u_0,h)$, 有 $\lambda_i=0$. 所以对 $h\in C(x_0)$, (PQ_h^3) 的对偶问题的可行集由 (非空) 的奇异 Lagrange 乘子集合与 0 构成, 因此 (PQ_h^3) 的对偶问题可以表示为

$$(\mathrm{DQ}_h^3)\qquad \max_{\lambda\in\Lambda^s(x_0,u_0)} D_xf(x_0,u_0)h+2D_uL^s(x_0,\lambda,u_0)d+D_{xx}^2L^s(x_0,\lambda,u_0)(h,h).$$

因为对偶问题(DQ_h^3) 是可行的, 所以具有和 (PQ_h^3) 相同的最优值.

注意到, 由引理 4.327, (DQ_h^3) 是有限值的充分必要条件是 (4.328) 成立, 其中

$$y := D_{xx}^2 G(x_0, u_0)(h, h) + 2D_u G(x_0, u_0)d.$$

由命题 2.41, 在 (4.328) 中取闭包是没有用的, 因此 (DQ_h^3) 是有限值的当且仅当 (PQ_h^3) 是可行的. 因为它们 (在其可行域上) 具有相同的目标函数, 实际上是相同的. 这当然同它们的值相等这一事实相符合.

对所有的临界方向取极小化, 得到问题

$$(\mathcal{DQ}^3) \qquad \min_{h \in C(x_0)} D_x f(x_0, u_0)h + \kappa(h), \tag{5.137}$$

其中

$$\kappa(h) := \sup_{\lambda \in \Lambda^s(x_0, u_0)} \{2D_u L^s(x_0, \lambda, u_0)d + D_{xx}^2 L^s(x_0, \lambda, u_0)(h, h)\}.$$

注意到, 因为 $L^s(x_0, \lambda, u_0)$ 关于 λ 是线性的, $\Lambda^s(x_0, u_0)$ 是锥, 函数 $\kappa(\cdot)$ 是满足

$$\sum_{i=1}^{p} \lambda_i [2D_u g_i(x_0, u_0)d + D_{xx}^2 g_i(x_0, u_0)(h, h)] \leqslant 0, \quad \forall \lambda \in \Lambda^s(x_0, u_0) \tag{5.138}$$

的向量 h 的集合的指示函数.

因此, 问题 $(\mathcal{DQ}^3)$ 可表示为在 $h \in C(x_0)$ 与约束 (5.138) 下极小化 $D_x f(x_0, u_0)h$.

定理 5.58　*设 x_0 是 (P_{u_0}) 的局部最优解. 设 $\Lambda(x_0, u_0)$ 是空集且沿方向 d 的 Gollan 条件成立, 则*

$$\limsup_{t \downarrow 0} \frac{v(u(t)) - v(u_0)}{t^{1/2}} \leqslant \frac{1}{2} \operatorname{val}(\mathcal{DQ}^3) < 0. \tag{5.139}$$

证明　第一不等式的证明类似于定理 5.56 的证明, 从略. 由上述讨论, 存在临界方向 h 满足 $D_x f(x_0, u_0)h < 0$. 另一方面, 易见, Gollan 条件可推出, 当 h 充分小时, 问题 (PQ_h^3) 是相容的. 证得第二不等式. □.

现在总结在 Lagrange 乘子集合是空集情况下相应的灵敏性结果.

定理 5.59　*设*

(i) 非扰动问题 (P_{u_0}) 具有唯一的最优解 x_0.

(ii) 沿方向 d 的 Gollan 条件成立.

(iii) 集合 $\Lambda(x_0, u_0)$ 是空集.

(iv) 二阶充分性条件 (5.104) 成立.

(v) 对于充分小的 $t > 0$, $(\mathrm{P}_{u(t)})$ 的可行集是非空的且一致有界的.

则对 $\mathcal{S}(\mathrm{P}_{u(t)})$ 的任何 $o(t)$ 最优解 $\bar{x}(t)$, 下述结论成立:

(a) $\bar{x}(t)$ 在 x_0 处是 Hölder 稳定的, 即 $\|\bar{x}(t) - x_0\| = O(t^{1/2})$.

(b) 最优值函数的下述展开式成立：

$$v(u(t)) = v(u_0) + t^{1/2}\,\mathrm{val}(\mathcal{D}Q^3) + o(t^{1/2}). \tag{5.140}$$

(c) $t^{-1/2}(\bar{x}(t) - x_0)$ 的任何极限点均是 $(\mathcal{D}Q^3)$ 的最优解.

证明 结论 (a) 在定理 5.57(a) 中已经证明. 用定理 5.58, 通过验证 $t^{-1/2}(\bar{x}(t) - x_0)$ 的任何极限点 $\bar{h}$ 均是 $(\mathcal{P}Q^3)$ 的解, 得到结论 (b) 与 (c). 因为与定理 5.57 相应的推证类似, 从略. □.

5.2.5 最优解与 Lagrange 乘子的高阶展开

这一节设 $f(x,u)$ 与 $G(x,u)$ 是 C^∞ 类的 (即无穷次可微), 且对给定的 $k \in \mathbb{N}$, $u(\cdot) : \mathbb{R}_+ \to U$ 是如下形式的 C^∞ 光滑映射 (路径)：

$$u(t) := u_0 + tu_1 + \cdots + \frac{t^k}{k!}u_k + o(t^k),$$

其中 $u_0, \cdots, u_k \in U$ 是某些固定的向量. 这一节讨论问题 $(\mathrm{P}_{u(t)})$ 的最优解与 Lagrange 乘子的高阶展开式. 为此目的, 用与 $(\mathrm{P}_{u(t)})$ 的最优性系统相联系的强正则性的概念 (见 5.1.3 节), 其中最优性系统可表示为下述广义方程的形式 (见 (5.47))：

$$Df(x,u) + D_xG(x,u)^*\lambda = 0, \quad G(x,u) \in N_K^{-1}(\lambda), \tag{5.141}$$

其中 $K := \{0\}_{\mathbb{R}^q} \times \mathbb{R}_-^{p-q}$. 回顾当 x 是 (P_u) 的局部最优解时, 在命题 5.38 中刻画的 (5.141) 解的强正则性.

定理 5.60 设 x_0 是 (P_{u_0}) 的局部最优解, 令 (x_0, λ_0) 是广义方程 (5.141) 的强正则解, 则存在 x_0 的邻域 $\mathcal{V}$, 满足对于充分小的 $t > 0$, $(\mathrm{P}_{u(t)})$ 具有唯一的局部最优解 $\bar{x}(t) \in \mathcal{V}$, 与之相联系有唯一的 Lagrange 乘子 $\overline{\lambda}(t)$, 映射 $t \to (\bar{x}(t), \overline{\lambda}(t))$ 是 C^∞ 类的.

证明 置 $I_+(\lambda) := \{i : \lambda_i > 0, i = q+1, \cdots, p\}$. 因为 λ_0 是与 x_0 相联系的唯一的 Lagrange 乘子, 定义 (5.116) 中的问题 $(\mathcal{D}Q)$ 可以表示为

$$\begin{aligned}
\min_{h_1 \in \mathbb{R}^n} \quad & D_uL(x_0, \lambda_0, u_0)u_2 + D^2L(x_0, \lambda, u_0)((h_1, u_1), (h_1, u_1)) \\
\text{s. t.} \quad & Dg_i(x_0, u_0)(h_1, u_1) = 0, \quad i \in \{1, \cdots, q\} \cup I_+(\lambda_0), \\
& Dg_i(x_0, u_0)(h_1, u_1) \leqslant 0, \quad i \in I(x_0) \setminus I_+(\lambda_0).
\end{aligned}$$

由于上述问题的二次目标函数的 Hesse 阵在 $(\mathcal{P}Q)$ 之可行集张成的子空间上是正定的, 约束是映上的, 这一问题有唯一的最优解 h_1, 与它相联系有唯一的 Lagrange 乘子 λ_1, 进一步, h_1 与 λ_1 满足

$$
\begin{aligned}
&D^2L(x_0,\lambda_0,u_0)(h_1,u_1)+DG(x_0,u_0)^*\lambda_1=0,\\
&\lambda_{1i}\geqslant 0,\quad \lambda_{1i}Dg_i(x_0,u_0)(h_1,u_1)=0,\quad i\in I(x_0)\setminus I_+(\lambda_0).
\end{aligned}
$$

于是得到

$$
\begin{aligned}
&D_xL(x_0+th_1,\lambda_0+t\lambda_1,u(t))=O(t^2),\\
&g_i(x_0+th_1,u(t))=O(t^2),\quad i\in\{1,\cdots,q\}\cup I_+(\lambda_0),\\
&g_i(x_0+th_1,u(t))\leqslant O(t^2),\quad i\in\{q+1,\cdots,p\},\\
&g_i(x_0+th_1,u(t))\geqslant O(t^2),\quad \text{若 } (\lambda_0+t\lambda_1)_i>0,\quad i\in\{q+1,\cdots,p\}.
\end{aligned}
$$

因此

$$
\begin{aligned}
&D_xL(x_0+th_1,\lambda_0+t\lambda_1,u(t))=O(t^2),\\
&G(x_0+th_1,u(t))+O(t^2)\in N_K^{-1}(\lambda_0+t\lambda_1).
\end{aligned}
\tag{5.142}
$$

根据定理 5.13, 由强正则性得到, 对 $t>0$ 充分小, $u=u(t)$ 时的最优性系统的局部的唯一解 $(\bar{x}(t),\overline{\lambda}(t))$ 满足 $(\bar{x}(t),\overline{\lambda}(t))=(x_0+th_1,\lambda_0+t\lambda_1)+O(t^2)$. 由此得到了 $\bar{x}(t)$ 与 $\overline{\lambda}(t)$ 的一阶展式.

现在示范如何计算二阶项. 其基本思想是忽视不属于由 (5.112) 定义的集合 $I_{u_1}(x_0,u_0,h_1)$ 的不等式约束, 将对应 $\lambda_0+t\lambda_1$ 的正分量的不等式转化为等式.

置

$$
I_+^1:=I_+(\lambda_0)\cup I_+(\lambda_1),
$$

计算下述系统解的二阶展式 $\bar{x}(t)=x_0+th_1+\dfrac{1}{2}t^2h_2+o(t^2)$, 这一系统对应于最优系统展式的二阶项:

$$
\begin{aligned}
&D^2_{(x,u)x}L(x_0,\lambda_0,u_0)(h_2,u_2)+D^3_{(x,u)(x,u)x}L(x_0,\lambda_0,u_0)((h_1,u_1),(h_1,u_1))\\
&\qquad+(D^2_{(x,u)x}G(x_0,u_0)(h_1,u_1))^*\lambda_1+DG(x_0,u_0)^*\lambda_2=0,\\
&D^2g_i(x_0,u_0)((h_1,u_1),(h_1,u_1))+Dg_i(x_0,u_0)(h_2,u_2)=0,\quad i\in\{1,\cdots,q\}\cup I_+^1,\\
&D^2g_i(x_0,u_0)((h_1,u_1),(h_1,u_1))+Dg_i(x_0,u_0)(h_2,u_2)\leqslant 0,\quad i\in I_{u_1}(x_0,u_0,h_1)\setminus I_+^1,\\
&\lambda_{2i}Dg_i(x_0,u_0)(h_2,u_2)=0,\quad i\in I_{u_1}(x_0,u_0,h_1)\setminus I_+^1.
\end{aligned}
$$

上述系统具有唯一解, 因为它是下述二次规划问题的最优性系统:

$$
\begin{aligned}
\min_{h_2}\quad & D^3_{(x,u)(x,u)(x,u)}L(x_0,\lambda_0,u_0)((h_1,u_1),(h_1,u_1);(h_2,u_2))\\
&\qquad+D^2_{(x,u)(x,u)}L(x_0,\lambda,u_0)((h_2,u_2),(h_2,u_2))\\
&\qquad+\lambda_1\cdot D^2_{(x,u)x}G(x_0,u_0)((h_1,u_1),h_2),\\
\text{s.t.}\quad & Dg_i(x_0,u_0)(h_2,u_2)=0,\quad i\in\{1,\cdots,q\}\cup I_+^1,\\
& Dg_i(x_0,u_0)(h_2,u_2)\leqslant 0,\quad i\in I_{u_1}(x_0,u_0,h_1)\setminus I_+^1,
\end{aligned}
$$

而上述问题的目标函数在由可行点的差张成的空间上是严格凸的, 且线性约束是映上的. 回到变分不等式的框架, 得到

$$DL\Big(x_0 + th_1 + \frac{1}{2}t^2h_2, \lambda_0 + t\lambda_1 + \frac{1}{2}t^2\lambda_2, u(t)\Big) = O(t^3),$$
$$G\Big(x_0 + th_1 + \frac{1}{2}t^2h_2, u(t)\Big) + O(t^3) \in N_K^{-1}\Big(\lambda_0 + t\lambda_1 + \frac{1}{2}t^2\lambda_2\Big). \quad (5.143)$$

由强正则性可得

$$(\bar{x}(t), \overline{\lambda}(t)) = \Big(x_0 + th_1 + \frac{1}{2}t^2h_2, \lambda_0 + t\lambda_1 + \frac{1}{2}t^2\lambda_2\Big) + O(t^3).$$

高阶情况的证明是类似的. □

现在证明, 在某些情况下, 子问题 $(\mathcal{PQ})$ 的最优解处非起作用的 $(\mathrm{P_{u(t)}})$ 的约束可以移掉而不会改变局部最优性. 因此可以在于 x_0 处起作用的约束的导数不是线性无关的情况下计算高阶展开. 这一定理在最优能量流问题的研究中将得到应用 (见 5.2.6 节).

定理 5.61 令 x_0 是 $(\mathrm{P_{u_0}})$ 的局部最优解. 设

(i) 问题 $(\mathcal{PQ})$ 具有唯一最优解 h_1.

(ii) 由 (5.112) 定义的线性化问题的起作用约束集 $\hat{I} := \{1, \cdots, q\} \cup I_{u_1}(x_0, u_0, h_1)$ 满足向量 $D_xg_i(x_0, u_0)$, $i \in \hat{I}$ 是线性无关的, 因此 (DL_{u_1}) 具有唯一的最优解 λ_0.

(iii) 对任何满足 $D_xg_i(x_0, u_0)h=0$, $i=1, \cdots, q$ 或 $\lambda_{0i}>0$ 的非零向量 $h\in\mathbb{R}^n$, 有

$$D_{xx}^2L(x_0, \lambda_0, u_0)(h, h) > 0,$$

则定理 5.60 的结论成立.

证明 考虑下述问题族

$$(\hat{\mathrm{P}}_u) \qquad \begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x, u), \\ \text{s. t.} \quad & g_i(x, u) = 0, \quad i = 1, \cdots, q, \\ & g_i(x, u) \leqslant 0, \quad i \in I_{u_1}(x_0, u_0, h_1). \end{aligned} \quad (5.144)$$

除了移掉不在 $I_{u_1}(x_0, u_0, h_1)$ 中的不等式约束, 这一问题与 (P_u) 是相同的. (问题 $(\hat{\mathrm{P}}_u)$) 在 x_0 处的起作用约束的导数是线性无关的, 二阶条件可推出 x_0 是 $(\hat{\mathrm{P}}_u)$ 的强正则解. 因此, 由定理 5.60 可得到问题 $(\hat{\mathrm{P}}_u)$ 局部最优解 $\hat{x}_u$ 的高阶展式. 因为 $\{q+1, \cdots, p\} \setminus I_{u_1}(x_0, u_0, h_1)$ 的约束在 $\hat{x}_u$ 处是不起作用的, 所以 $\hat{x}_u$ 也是 (P_u) 的局部最优解. 由定理 5.60 可得结论. □

5.2.6 电子网络

我们来描述 Lipschitz 展开理论在所谓的直流最优能量流问题中的应用. 考虑标号从 0 到 n 的 $n+1$ 个结点, V_k 是结点 k 处的电压. 结点 0 起特殊的作用. 因为

参考电压 V_0 是该问题的基准数据, 令 $Y_{kl} \geqslant 0$ 是结点 k 与 l 间的导纳 (即电阻的倒数), I_{kl} 是结点 k 到结点 l 的电流. 这一问题是在遵守 Ohm 与 Kirchhoff 定律及满足电压与结点处输入能量的界值约束的前提下极小化整个网络能量的损失.

Ohm 定律是 $I_{kl} = Y_{kl}(V_k - V_l), 0 \leqslant k \neq l \leqslant n$, 其中 Y 是 $n+1$ 阶的对称方阵, 其对角线元素都是零, 只具有非负的元素. Y_{kl} 的零值意味着 k 与 l 结点间没有物理连接. 设该网络是连通的, 其含义是任何两结点均可由具有正值 Y_k 的弧构成的路径连接. Kirchhoff 定律是 $J_k = \sum_{l=0}^{n} I_{kl}$, 其中 J_k 是结点 k 处的电流的输入. 结点 k 处的能量输入是 $P_k := J_k V_k$. 令 Z 是如下定义的 $(n+1) \times (n+1)$ 矩阵:

$$Z_{kk} := \sum_{l=0}^{n} Y_{kl}; \quad Z_{kl} := -Y_{kl}, \quad 0 \leqslant k \neq l \leqslant n.$$

因为 $V^{\mathrm{T}} Z V = \sum_{0 \leqslant k < l \leqslant n} Y_{kl}(V_k - V_l)^2$, 矩阵 Z 是正半定的, 且相联系的二次型表示结点间的连接上的电能的总损失. 设每一结点处的电压均是正的, 由前述关系可推出能量方程:

$$(ZV)_k = P_k / V_k, \quad k = 1, \cdots, n \tag{5.145}$$

(因为 V_0 是基准数据, 不考虑结点 0 处的能量方程). 考虑在电压与能量输入的界值约束之下, 极小化所有连接上的能量损失总和的问题:

$$(\mathcal{P}_\varepsilon) \qquad \begin{aligned} \min_{V,P} \quad & \frac{1}{2} V^{\mathrm{T}} Z V, \\ \text{s. t.} \quad & (5.145); \\ & V_0 = \varepsilon^{-\frac{1}{2}}; \\ & \varepsilon^{-\frac{1}{2}} V^{b} \leqslant V \leqslant \varepsilon^{-\frac{1}{2}} V^{\sharp}; \\ & P^{b} \leqslant P \leqslant P^{\sharp}, \end{aligned}$$

这里 V^b, $V^\sharp$, P^b 与 $P^\sharp$ 是满足 $V_k^b < V_k^\sharp$ 与 $P_k^b < P_k^\sharp$, $k = 1, \cdots, n$ 的 n 维向量. V 的界约束可理解为对指标 1 到 n 成立. 最后, $\varepsilon > 0$ 与虚拟电压 V_0 值的数量量级相联系. 为计算当 ε 接近于 0 的上述问题的最优解的渐进展开, 做下述的变量的尺度化是合适的:

$$\widetilde{V} := \varepsilon^{\frac{1}{2}} V, \quad \widetilde{P} := \varepsilon P.$$

因为目标函数是 2 度正齐次的, 在此变化之下, 能量方程与最优化问题可以表示为

$$(Z\widetilde{V})_k = \widetilde{P}_k / \widetilde{V}_k, \quad k = 1, \cdots, n, \tag{5.146}$$

$$(\widetilde{\mathcal{P}}_\varepsilon) \qquad \begin{aligned} \min_{\widetilde{V},\ \widetilde{P}} \quad & \frac{1}{2}\,\widetilde{V}^{\mathrm{T}} Z \widetilde{V}, \\ \text{s. t.} \quad & (5.146); \\ & V^b \leqslant \widetilde{V} \leqslant V^\sharp; \\ & \varepsilon P^b \leqslant \widetilde{P} \leqslant \varepsilon P^\sharp, \end{aligned}$$

其中 $\widetilde{V}_0$ 可以置为 1, 可视为给定的参数. 从数学的观点看, 这一变换表明, 它等价于或者令参考电压趋于无穷大而使能量输入保持在固定的范围或者让能量输入趋于零而保持固定的参考电压. 换言之, 高的电压近似不是别的, 只是小的能量输入的近似. 置 ε 趋于 0, 可以得到下述的极限问题

$$(\widetilde{\mathcal{P}}_0) \qquad \begin{aligned} \min_{\widetilde{V},\ \widetilde{P}} \quad & \frac{1}{2}\,\widetilde{V}^{\mathrm{T}} Z \widetilde{V}, \\ \text{s. t.} \quad & (5.146); \\ & V^b \leqslant \widetilde{V} \leqslant V^\sharp; \\ & 0 \leqslant \widetilde{P} \leqslant 0. \end{aligned}$$

设计高电压网络的物理动机是减小连接上的能量的损失. 为达此目的, 加在电压上的界值约束一定要允许可接近于参考值, 从而使假设 $V^b \leqslant \mathbf{1} \leqslant V^\sharp$ 成为是有意义的, 其中 $\mathbf{1}$ 是 $\mathbb{R}^n$ 中所有分量均是 1 的向量.

引理 5.62 令 $V^b \leqslant \mathbf{1} \leqslant V^\sharp$, 则极限问题 $(\widetilde{\mathcal{P}}_0)$ 具有唯一的最优解 $(\widetilde{V}^0, \widetilde{P}^0)$, 定义为 $\widetilde{V}^0 = \mathbf{1}$, $\widetilde{P}^0 = 0$.

证明 先验证上述定义的 $(\widetilde{V}^0, \widetilde{P}^0)$ 是 $(\widetilde{\mathcal{P}}_0)$ 的可行点. 因为在整个网络上 $\widetilde{V}^0$ 是常值, 所以 $Z\widetilde{V}^0 = 0$. 另外, $\widetilde{V}^0 = \mathbf{1} > 0$ 可推出 $\widetilde{P}^0/\widetilde{V}^0$ 是有定义的, 具有值 0. 因此, 能量方程是成立的, 界值约束也是成立的.

由于 $(\widetilde{V}^0, \widetilde{P}^0)$ 联系着零的目标函数, 目标函数是非负的, 从而 $(\widetilde{V}^0, \widetilde{P}^0)$ 是 $(\widetilde{\mathcal{P}}_0)$ 的最优解. 任何其他的最优解 $(\widetilde{V}, \widetilde{P})$ 也联系着零的目标函数, 因此必满足 $(Z\widetilde{V}^0)_k = 0$, $k = 1, \cdots, n$. 于是有 $\widetilde{V} = \widetilde{V}^0$, 由能量方程也可推出 $\widetilde{P} = \widetilde{P}^0$. □

极限问题不满足 Mangasarian-Fromovitz 约束规范条件, 因为不等式的线性化不能严格地被满足. 对电压的起作用界约束, 需要下述记号: $\overline{V}^b$, $\overline{V}^\sharp$ 是如下定义的 n 维向量:

$$\overline{V}^b_k = \begin{cases} V^b_k, & \text{若} V^b_k = 1, \\ -\infty, & \text{否则}, \end{cases} \qquad \overline{V}^b_\sharp = \begin{cases} V^\sharp_k, & \text{若} V^\sharp_k = 1, \\ +\infty, & \text{否则}. \end{cases}$$

令 Z^R 是由 Z 删除第一行与第一列得到的. 因为网络是连通的, Z^R 是可逆的. 在 $(\widetilde{V}^0, \widetilde{P}^0, 0)$ 处关于 $(\widetilde{V}, \widetilde{P}, \varepsilon)$ 线性化 $(\widetilde{\mathcal{P}}_\varepsilon)$ 的数据, 用到等式 $(\widetilde{V}^0)^{\mathrm{T}} Z = 0$, 得到线性化

问题的表达式

$$(\mathrm{LP})\qquad \begin{aligned} &\min_{\mathrm{d}V,\ \mathrm{d}P} && 0,\\ &\text{s. t.} && Z^R\mathrm{d}V=\mathrm{d}P,\quad \overline{V}^b\leqslant \mathbf{1}+\mathrm{d}V\leqslant \overline{V}^\sharp;\\ &&& P^b\leqslant \mathrm{d}P\leqslant P^\sharp. \end{aligned}$$

因此, $\mathcal{S}(\mathrm{LP})=F(\mathrm{LP})$, 其中 $F(\mathrm{LP})$ 是 (LP) 的可行集. 例如, 如果 $V_k^b<\mathbf{1}<V_k^\sharp$ 对所有 $k=1,\cdots,n$ 成立, 则 Gollan 条件成立.

引理 5.63　设 $V^b\leqslant \mathbf{1}\leqslant V^\sharp$. 若极限问题满足 Gollan 条件, 则 (LP) 的对偶问题具有唯一的最优解, 即零乘子.

证明　由于问题 $\widetilde{\mathcal{P}}_0$ 的目标函数在极限问题的最优解处的导数是零, Lagrange 乘子集合是锥. (LP) 的对偶问题是在此锥上极大化线性的目标函数, 其最优解集本身也是锥. 由于 Gollan 条件成立, 这个集合是非空有界的, 因而退化为 $\{0\}$. □

在此以后设 Gollan 条件是成立的. 由于 (LP) 的对偶问题的最优解集是 $\{0\}$, (5.116) 定义的二阶子问题 $(\mathcal{DQ})$ 的目标函数在应用中简化为目标函数的 Hesse 形式. 由 $\mathcal{S}(\mathrm{LP})=F(\mathrm{LP})$, 可将此问题表示为

$$(\mathcal{PQ})\qquad \begin{aligned} &\min_{\mathrm{d}V,\ \mathrm{d}P} && (\mathrm{d}V)^{\mathrm{T}}Z^R\mathrm{d}V,\\ &\text{s. t.} && (\mathrm{d}V,\ \mathrm{d}P)\in F(\mathrm{LP}), \end{aligned}$$

或等价地,

$$\begin{aligned} &\min_{\mathrm{d}P} && (\mathrm{d}P)^{\mathrm{T}}(Z^R)^{-1}\mathrm{d}P,\\ &\text{s. t.} && \overline{V}^b\leqslant \mathbf{1}+(Z^R)^{-1}\mathrm{d}P\leqslant \overline{V}^\sharp;\\ &&& P^b\leqslant \mathrm{d}P\leqslant P^\sharp. \end{aligned}$$

这一问题具有唯一的最优解 $(\overline{\mathrm{d}V},\ \overline{\mathrm{d}P})$, 即原点沿与 $(Z^R)^{-1}$ 相联系的范数下到可行集的投影. 下述引理表明标准的二阶条件成立.

引理 5.64　考虑极限问题 $(\widetilde{\mathcal{P}}_0)$. 只要 $(\mathrm{d}V,\ \mathrm{d}P)\neq 0$ 属于线性化等式约束的零空间, 即只要 $Z^R\mathrm{d}V=\mathrm{d}P$, 就有 $(\mathrm{d}V)^{\mathrm{T}}Z^R\mathrm{d}V>0$.

证明　由于二次型 $(\mathrm{d}V)^{\mathrm{T}}Z^R\mathrm{d}V$ 是正定的, 有 $(\mathrm{d}V)^{\mathrm{T}}Z^R\mathrm{d}V=0$ 当且仅当 $\mathrm{d}V=0$, 这表明 $\mathrm{d}P=Z^R\mathrm{d}V$ 也是零. 从而证得结论. □

定理 5.65　设 $V^b\leqslant \mathbf{1}\leqslant V^\sharp$, 问题 $(\widetilde{\mathcal{P}}_0)$ 满足 Gollan 条件, 则对充分小的 ε, $(\widetilde{\mathcal{P}}_\varepsilon)$ 具有唯一的最优解 $(\widetilde{V}^\varepsilon,\widetilde{P}^\varepsilon)$, 且下述表达式成立:

$$\begin{aligned} \mathrm{val}(\widetilde{\mathcal{P}}_\varepsilon)&=\frac{1}{2}\varepsilon^2(\overline{\mathrm{d}V})^{\mathrm{T}}Z^R\overline{\mathrm{d}V}+o(\varepsilon^2),\\ (\widetilde{V}^\varepsilon,\widetilde{P}^\varepsilon)&=(\widetilde{V}^0,\widetilde{P}^0)+\varepsilon(\overline{\mathrm{d}V},\overline{\mathrm{d}P})+o(\varepsilon). \end{aligned}$$

与 $(\widetilde{V}^{\varepsilon}, \widetilde{P}^{\varepsilon})$ 相联系的是非空的一致有界的乘子集合 Λ_{ε}. 乘子集在下述定义下收敛到 0, 即若 $\varepsilon \downarrow 0$, $\lambda^{\varepsilon} \in \Lambda_{\varepsilon}$, 则 $\lambda^{\varepsilon} \to 0$.

证明 显然, 对充分小的 ε, $(\widetilde{\mathcal{P}}_{\varepsilon})$ 的可行集合是非空的一致有界的. 由定理 5.53, 用引理 5.62~5.64 可得到结论. □

返回到非尺度化的变量, 得到能量损失是

$$\frac{1}{2}\varepsilon(\overline{\mathrm{d}V})^{\mathrm{T}} Z^{R}(\overline{\mathrm{d}V}) + o(\varepsilon),$$

这意味着, 对这一模型, 能量损失是参考电压的逆的平方阶的.

现在研究在较强的假设下 $(\widetilde{\mathcal{P}}_{\varepsilon})$ 的最优解与相应乘子的高阶展式.

引理 5.66 设问题 (LP) 满足在网络的节点上 $\overline{\mathrm{d}V}$ 与 $\overline{\mathrm{d}P}$ 的界值约束总不是同时起作用的, 则 (等式与不等式) 起作用约束的导数是线性无关的.

证明 设关于电压 (或能量输入) 的界约束在 $\mathcal{S}_V^a$(或 $\mathcal{S}_P^a$) 上是起作用的. 由假设有 $\mathcal{S}_V^a \cap \mathcal{S}_P^a = \varnothing$. 置于 $\mathcal{S} \setminus (\mathcal{S}_V^a \cup \mathcal{S}_P^a)$ 上 $\mathrm{d}P$ 为 0. 必须检验可以求解在 $\mathcal{S}_V^a$ 与参考节点上电压固定, 其他结点上电流输入是固定的线性 DC 问题. 因为网络是连通的, 所以这一问题具有唯一解. □

定理 5.67 设

(i) $V^b \leqslant \mathbf{1} \leqslant V^{\sharp}$.

(ii) 于网络的结点处关于 $\overline{\mathrm{d}V}$ 与 $\overline{\mathrm{d}P}$ 的界约束总不会同时起作用.

则映射 $\varepsilon \mapsto (\widetilde{V}^{\varepsilon}, \widetilde{P}^{\varepsilon}, \lambda^{\varepsilon}, \mathrm{val}(\widetilde{\mathcal{P}}_{\varepsilon}))$, 其中 $(\widetilde{V}^{\varepsilon}, \widetilde{P}^{\varepsilon}) \in \mathcal{S}(\widetilde{\mathcal{P}}_{\varepsilon})$ 且 λ^{ε} 是与能量方程相联系的乘子, 是有定义的且对充分小的 ε 是实解析的.

证明 本定理的假设与引理 5.66 表明, 定理 5.61 的假设是满足的. 证毕. □

5.2.7 悬链问题

现在介绍静态力学中的一个简单问题, 即悬链或悬链线问题, 这一问题将说明具有 Lagrange 乘子或不具有 Lagrange 乘子的 Hölder 展开理论的作用. 考虑具有 m 个刚性连接的链 $(m > 1)$, 这些连接从 1 标记到 m, 每一连接具有单位重量, 具有固定的端点. 考虑计算这一链的平衡位置. 这一问题的正规描述如下: 用 (y_k, z_k) 记沿从 1 到 m 的连接的水平与竖直的增量, 设一端点在 (0,0) 处固定, 另一端点的位置是 $\sum_{i=1}^{m}(y_i, z_i)$, 受约束为 $(y^e(u), z^e(u))$, 其中 $u \in \mathbb{R}$ 可视为扰动的参数. 势能的表达式是

$$E(y,z) = \frac{1}{2}\sum_{k=1}^{m}\Big(\sum_{i=1}^{k} z_i + \sum_{i=1}^{k-1} z_i\Big) = \sum_{k=1}^{m} \alpha_k z_k, \tag{5.147}$$

其中 $\alpha_k = m - k + \frac{1}{2}$, $k = 1, \cdots, m$. 悬链的平衡位置可以通过在所涉及的物理约

束下极小化势能得到, 这导致下述最优化问题

$$\begin{aligned}\min_{y,z}\quad & E(y,z),\\ \text{s. t.}\quad & y_k^2+z_k^2-1=0,\ k=1,\cdots,m,\\ & \sum_{k=1}^m (y_k,z_k)=(y^e(u),z^e(u)).\end{aligned} \tag{5.148}$$

设对 $u_0:=0$, 两端点间的距离恰好是 m, 则非扰动问题有唯一的可行点 (悬链是直线), 因而是唯一的最优解. 考虑两种情况, 其中悬链对 $u=u_0=0$ 或者是竖直的或者是水平的, 即设或者 $y^e(u)=0,\ z^e(u)=m-u$ 或者 $y^e(u)=m-u,\ z^e(u)=0$, 令其中一端点朝另一端点前进一小步, 即 u 从零增加到小的正数, 在后面置 $x:=(y,z)$.

1. 竖直悬链的扰动

置 $y_0=0$, 令 z_0 是 $\mathbb{R}^m$ 中分量均是 1 的向量, 有 $x_0:=(y_0,z_0)$ 是非扰动问题的唯一的最优解. 容易验证, 如果将最后的等式变成不等式, 悬链的位置是相同的, 因而得到下述最优化问题:

$$\begin{aligned}\min_{y,z}\quad & E(y,z),\\ \text{s. t.}\quad & \frac{1}{2}y_k^2+\frac{1}{2}z_k^2-\frac{1}{2}=0,\ k=1,\cdots,m,\\ & \sum_{k=1}^m y_k=0,\\ & m-u-\sum_{k=1}^m z_k\leqslant 0.\end{aligned} \tag{5.149}$$

因为 u 是数值, 可取 $d=1$, 因此这里 $t=u$. 现在验证, 问题 (5.149) 显然满足 Gollan 条件, 尽管问题 (5.148) 不满足. 事实上, 线性化的等式约束是映上的, 因为对任何 (γ,β), 线性系统

$$h_k^z=\gamma_k,\quad k=1,\cdots,m;\quad \sum_{k=1}^m h_k^y=\beta;\quad \gamma\in\mathbb{R}^m,\quad \beta\in\mathbb{R} \tag{5.150}$$

显然有最优解. 若 $(\gamma,\beta)=0$, 则 $h_k^z=0$. 于是得到线性化等式约束的零空间内的变化量不会严格地满足线性化起作用的不等式约束 $-\sum\limits_k h_k^z\leqslant 0$. 所以 Mangasarian-Fromovitz 约束规范是不成立的. 另一方面, 由于 (5.111) 对 $h=(h^y,h^z)=0$ 是成立的, 所以 Gollan 条件成立.

Lagrange 函数 $L(y,z,\lambda,u)$ 的表达式是

$$\sum_{k=1}^{m}\alpha_k z_k+\sum_{k=1}^{m}\lambda_k\Big(\frac{1}{2}y_k^2+\frac{1}{2}z_k^2-\frac{1}{2}\Big)+\lambda_{m+1}\sum_{k=1}^{m}y_k+\lambda_{m+2}\Big(m-u-\sum_{k=1}^{m}z_k\Big).$$

当 $\lambda_{m+1}=0$ 且 $\alpha_k+\lambda_k-\lambda_{m+2}=0,\ k=1,\cdots,m$ 时, Lagrange 函数关于 (y,z) 的导数是零. 所以,Lagrange 乘子集合是

$$\Lambda(x_0)=\{\lambda\in\mathbb{R}^{m+2}:\lambda_{m+1}=0;\ \lambda_{m+2}\geqslant 0;\ \lambda_k=\lambda_{m+2}-\alpha_k,\ 1\leqslant k\leqslant m\}.$$

临界锥的方程是

$$\sum_{k=1}^{m}\alpha_k h_k^z\leqslant 0;\quad h_k^z=0,\quad k=1,\cdots,m;\quad \sum_{k=1}^{m}h_k^y=0;\quad -\sum_{k=1}^{m}h_k^z\leqslant 0,$$

即

$$C(x_0)=\Big\{(h^y,h^z):\ h^z=0\ 且\ \sum_{k=1}^{m}h_k^y=0\Big\}.$$

现在验证二阶充分条件 (5.104). 给定临界方向 h, 用到 $h^z=0$, 计算

$$D_{xx}^2L(x_0,\lambda,0)(h,h)=\sum_{k=1}^{m}\lambda_k(h_k^y)^2.$$

若 λ_{m+2} 充分大, 则 $\lambda_k\geqslant 1,\ k=1,\cdots,m$. 此种情况下, 当 h 是非零的临界方向时, Lagrange 函数的二阶变化是非零的, 即二阶充分性条件成立 (可以验证, 强方向二阶充分条件不成立).

还可以验证, 对正的 $t>0$, 集合 $\mathcal{S}(\mathrm{P}_{u(t)})$ 是非空的一致有界的. 所以, 定理 5.57 的假设是成立的. 问题 $(\mathcal{D}Q^2)$ 可表示为

$$\min_{h\in C(x_0)}\ \sup_{\lambda\in\Lambda(x_0)}\ \Delta(h,\lambda),$$

其中

$$\Delta(h,\lambda):=-2\lambda_{m+2}+\sum_{k=1}^{m}\lambda_k(h_k^y)^2=\lambda_{m+2}\Big(\sum_{k=1}^{m}(h_k^y)^2-2\Big)-\sum_{k=1}^{m}\alpha_k(h_k^y)^2.$$

若 $\sum_k(h_k^y)^2>2$, 则 $\sup\{\Delta(h,\lambda);\lambda\in\Lambda(x_0)\}=+\infty$. 否则, $\lambda_{m+2}=0$ 时上确界取到. 因此, $(\mathcal{D}Q^2)$ 与下述问题具有相同的解集

$$\begin{aligned}&\min_{h\in C(x_0)}\ -\frac{1}{2}\sum_{k=1}^{m}\alpha_k(h_k^y)^2,\\ &\text{s. t.}\ \sum_{k=1}^{m}(h_k^y)^2\leqslant 2.\end{aligned}\tag{5.151}$$

用临界锥的表达式将此问题表示为

$$\begin{aligned}\min_{h^y} \quad & -\frac{1}{2}\sum_{k=1}^{m}\alpha_k(h_k^y)^2,\\ \text{s. t.} \quad & \frac{1}{2}\sum_{k=1}^{m}(h_k^y)^2-1\leqslant 0, \quad \sum_{k=1}^{m}h_k^y=0.\end{aligned}$$

问题具有最优解 (紧致的非空的可行集)、凹目标函数与凸的约束, 约束在每一可行点处均是正则的. 与最优解 h^y 相联系的乘子 (η,ν), $\eta\geqslant 0$ 满足

$$-\alpha_k h_k^y+\eta h_k^y+\nu=0, \quad k=1,\cdots,m. \tag{5.152}$$

下证 $\nu\neq 0$. 若 $\nu=0$, 则 $(\eta-\alpha_k)h_k^y=0$, 对所有 k 成立. 由于 α_k 的所有分量具有不同的值, 得到 $h^y=0$, 除非只有一个分量. 于是, 等式约束可推出 $h^y=0$. 因此, 由于不等式约束是非起作用的, $\eta=0$. 然而, 在 $h^y=0$ 处二阶必要条件不成立. 证得 $\nu\neq 0$. 由 (5.152) 可得

$$h_k^y=\nu/(\alpha_k-\eta); \quad k=1,\cdots,m. \tag{5.153}$$

对指标 k 求和, 得到

$$0=\sum_{k=1}^{m}h_k^y=\nu\sum_{k=1}^{m}\frac{1}{\alpha_k-\eta}. \tag{5.154}$$

由于 $\nu\neq 0$, 可给出关于数值未知数 η 的数值方程:

$$\sum_{k=1}^{m}\frac{1}{\alpha_k-\eta}=0.$$

研究这一方程. α_k 是严格递减的序列, 若 $\eta<\alpha_m$ 或 $\eta>\alpha_1$, 则和式所有的项均是同号的. 因此方程的解必属于 (α_m,α_1), 且不能等于某一 α_k. 在 $\{1,\cdots,m-1\}$ 中固定 k, 研究 $\eta\in(\alpha_{k+1},\alpha_k)$ 这样的解, 方程的左端从 $-\infty$ 严格地递增到 $+\infty$, 于是在每一区间 (α_{k+1},α_k) 内, (5.154) 均有唯一解, 记为 η_k.

通过将由 (5.153) 给出的 h^y 的表达式插入到起作用的不等式约束中确定与 η_k 相联系的乘子 $\nu=\nu_k$ 的值, 可得到

$$\nu_k=\pm\sqrt{2\Big/\sum_{k=1}^{m}(\alpha_k-\eta_k)^{-2}}.$$

由 (5.153), 与 ν 的两个值联系着 h^y 的互为相反的值, 记为 $\pm h^{y,k}$, 它们生成目标函数的相等的值. 由于这一问题的对称性, 这是很自然的. 对应于乘子 η_k 的 $(\mathcal{D}Q^2)$ 的

最优系统的互为相反的解, $\pm h^{y,k}$, 从 1 到 k 的分量均具有相同的符号, 因而它们具有相反的符号.

(5.151) 的最优解 h^y 的分量必是非增的, 因为否则的话, 对 h^y 的分量做置换, 使它们关于 k 是递减的, 则可得到另一可行点, 它联系的目标函数值更小. 于是有 $\pm h_1^y$ 是 $(\mathcal{DQ}^2)$ 的两个最优解. 结果, 比如 h_1^y 是负的, 所有 h^y 的其他分量均是正的, 且为严格递减的 (由 (5.153)). 因为子问题具有两个最优解, 定理 5.57 推出, $u \geqslant 0$ 的任何 $o(u)$ 最优路径具有形式 $x_0 + s(u)\sqrt{u}h + o(u^{\frac{1}{2}})$, 其中 $s(u)$ 在 $\{-1, 1\}$ 中取值.

2. 水平悬链的扰动

不难看到, 通过将最后的等式变化为不等式, 可以得到问题 (5.148) 的等价形式, 即问题 (5.148) 等价于下述最优化问题:

$$
\begin{aligned}
\min_{y,z} \quad & E(y,z), \\
\text{s. t.} \quad & \frac{1}{2}y_k^2 + \frac{1}{2}z_k^2 - \frac{1}{2} = 0, \ k = 1, \cdots, m, \\
& \sum_{k=1}^m z_k = 0, \\
& m - u - \sum_{k=1}^m y_k \leqslant 0.
\end{aligned} \tag{5.155}
$$

再次可取 $d = 1$, $t = u$. 类似于竖直悬链情况用到的推证, 验证问题 (5.149) 满足 Gollan 条件.

广义 Lagrange 函数 $L^g(y, z, \lambda, u)$ 的表达式为

$$
\lambda_0 \sum_{k=1}^m \alpha_k z_k + \sum_{k=1}^m \lambda_k \Big(\frac{1}{2}y_k^2 + \frac{1}{2}z_k^2 - \frac{1}{2}\Big) + \lambda_{m+1} \sum_{k=1}^m z_k + \lambda_{m+2}\Big(m - u - \sum_{k=1}^m y_k\Big).
$$

当

$$
\lambda_k - \lambda_{m+2} = 0, \quad \lambda_0 \alpha_k + \lambda_{m+1} = 0, \quad k = 1, \cdots, m
$$

时, Lagrange 函数关于 (y, z) 的导数是零向量. 因此, Lagrange 乘子集合是空集, 而广义 Lagrange 乘子的集合是

$$
\Lambda^g(x_0) = \{(\lambda_0, \lambda) \in \mathbb{R}^{m+2} : \lambda_0 = \lambda_{m+1} = 0;\ \lambda_k = \lambda_{m+2} > 0,\ 1 \leqslant k \leqslant m\}.
$$

临界锥是

$$
C(x_0) = \Big\{(h^y, h^z) : h^y = 0,\ \sum_{k=1}^m \alpha_k h_k^z \leqslant 0,\ \sum_{k=1}^m h_k^z = 0\Big\}.
$$

检验二阶充分性条件 (5.104). 给定临界方向 h, 用到 $h^y = 0$, 得到[①]

$$D_{xx}^2 L(x_0, \lambda, 0)(h, h) = \sum_{k=1}^{m} \lambda_k (h_k^z)^2$$

当 h 是非零临界方向时, 这个值是正数, 因此 (奇异的) 二阶充分性条件成立. 因为对正的 t, 集合 $\mathcal{S}(\mathrm{P}_{u(t)})$ 是非空的一致有界的, 定理 5.59 的假设成立. 子问题 $(\mathcal{DQ}^3)$ 是

$$\min_{h \in C_0} \sup_{\lambda \in \Lambda^g(x_0)} \Delta(h, \lambda),$$

其中

$$\Delta(h, \lambda) := \sum_{k=1}^{m} \alpha_k h_k^z + \lambda_{m+2} \left(\sum_{k=1}^{m} (h_k^z)^2 - 2 \right).$$

若 $\sum_k (h_k^z)^2 > 2$, 则 $\sup\{\Delta(h, \lambda) : \ \lambda \in \Lambda(x_0)\} = +\infty$. 否则, 上确界在 $\lambda_{m+2} = 0$ 时取到. 注意到 C_0 的表达式, 得到 $(\mathcal{DQ}^3)$ 与下述问题具有相同的最优解集:

$$\begin{aligned} \min_{h^z} \quad & \sum_{k=1}^{m} \alpha_k h_k^z, \\ \text{s. t.} \quad & \frac{1}{2} \sum_{k=1}^{m} (h_k^z)^2 - 1 \leqslant 0, \\ & \sum_{k=1}^{m} h_k^z = 0. \end{aligned}$$

这是规范的凸问题, 它的最优解由连接的末端点的高度变化更好地表达出来. 第 k 条连接的末端点的高度是 $\sum_{i=1}^{k} z_i$. 相联系的增量 $\bar{h}_k := \sum_{i=1}^{k} h_i^z$ 是下述问题的解:

$$\begin{aligned} \min_{\bar{h}} \quad & \sum_{k=1}^{m-1} \bar{h}_k, \\ \text{s. t.} \quad & \frac{1}{2} \sum_{k=1}^{m} (\bar{h}_k - \bar{h}_{k-1})^2 - 1 \leqslant 0, \end{aligned}$$

其中置 $\bar{h}_0 = \bar{h}_m = 0$. 令 α 是与这一约束相联系的 Lagrange 乘子, 则 $\alpha > 0$, $\bar{h}$ 是下述问题的最优解:

$$\min \sum_{k=1}^{m-1} \bar{h}_k + \frac{\alpha}{2} \sum_{k=1}^{m} (\bar{h}_k - \bar{h}_{k-1})^2.$$

① 应为 $D_{xx}^2 L^g(x_0, \lambda, 0)(h, h)$.

因此, 可如下计算 $\bar{h}$: 令 $\hat{h}$ 是下述线性系统的解:

$$\hat{h}_{k+1} - 2\hat{h}_k + \hat{h}_{k-1} = 1; \quad \hat{h}_0 = \hat{h}_m = 0.$$

注意到, 这一线性系统类似于将一维的具有 Dirichlet 边界条件与常值右端项的 Poisson 方程离散化得到的系统, 计算相联系的能量 $\eta := \frac{1}{2}\sum_{k=1}^{m}(\hat{h}_k - \hat{h}_{k-1})^2$, 则 $\bar{h} := \eta^{-\frac{1}{2}}\hat{h}$ 是 $(\mathcal{D}\mathcal{Q}^3)$ 的唯一最优解. 令 h^z 是相应的 z 分量增量, 则由定理 5.59 得

$$x(u) = x(0) + u^{\frac{1}{2}}(0, h^z) + o(u^{\frac{1}{2}}).$$

5.3 半定规划

这一节考虑下述形式的最优化问题:

$$\text{(P)} \qquad \begin{aligned} \min_{x \in Q} \quad & f(x), \\ \text{s. t.} \quad & G(x) \preceq 0, \end{aligned} \tag{5.156}$$

其中 Q 是欧氏空间 $\mathbb{R}^n$ 的非空凸闭子集, $f : \mathbb{R}^n \to \mathbb{R}$, $G : \mathbb{R}^n \to \mathcal{S}^p$ 是从 $\mathbb{R}^n$ 到 $p \times p$ 对称矩阵空间 $\mathcal{S}^p$ 的映射. 记号 $A \preceq 0 (A \succeq 0)$ 意味着对称阵 A 是负半定的 (正半定的). 这一节假设 $f(x)$ 与 $G(x)$ 是连续的. 注意到, 若 $G(x)$ 是可微的, 则在点 $x \in \mathbb{R}^n$ 处的微分可以写成形式 $DG(x)h = \sum_{i=1}^{n} h_i G_i(x)$, 其中 $G_i(x) := \frac{\partial G(x)}{\partial x_i}$, $i = 1, \cdots, n$ 是 $G(x)$ 的元素关于 x_i 的偏导数构成的 $p \times p$ 矩阵.

称问题 (5.156) 为 (非线性的) 半定规划 (SDP) 问题. 尤其, 若 $Q = \mathbb{R}^n$, 目标函数是线性的, 即 $f(x) := \sum_{i=1}^{n} c_i x_i$, 约束映射是仿射的, 即 $G(x) := A_0 + \sum_{i=1}^{n} x_i A_i$, 其中 $A_0, A_1, \cdots, A_n \in \mathcal{S}^p$ 是给定的矩阵, 问题 (5.156) 成为如下的线性的半定规划问题:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \sum_{i=1}^{n} c_i x_i \\ \text{s. t.} \quad & A_0 + \sum_{i=1}^{n} x_i A_i \preceq 0. \end{aligned} \tag{5.157}$$

问题 (5.156) 可以表示为下述形式:

$$\begin{aligned} \min_{x \in Q} \quad & f(x), \\ \text{s. t.} \quad & G(x) \in K, \end{aligned} \tag{5.158}$$

其中 $K := \mathcal{S}_-^p$ 是负半定的 $p \times p$ 对称矩阵构成的锥, 即问题 (5.156) 的可行集可以定义为具有形式 $\{x \in Q : G(x) \in K\}$ 的 "锥约束".

用 $x^{\mathrm{T}}(A^{\mathrm{T}})$ 记列向量 x(矩阵 A) 的转置. 对 $k \times l$ 矩阵, 用 $\mathrm{vec}(A)$ 记由 A 的列拉直得到的 $kl \times 1$ 向量. 对两个矩阵, 一个 $k \times l$ 矩阵 A 与一个 $m \times n$ 矩阵 B, 用 $A \otimes B$ 记它们的 Kronecker 积, 即 $A \otimes B$ 是如下定义的 (块形式的)$km \times ln$ 矩阵:

$$A \otimes B := \begin{bmatrix} a_{11}B & \cdots & a_{1l}B \\ \vdots & & \vdots \\ a_{k1}B & \cdots & a_{kl}B \end{bmatrix}.$$

Kronecker 乘积有如下的一些基本性质 (假设相应的运算是容许的):

$$\begin{aligned} (A \otimes B)(C \otimes D) &= (AC) \otimes (BD), \\ (A \otimes B)^{-1} &= A^{-1} \otimes B^{-1}, \\ \mathrm{vec}(ABC) &= (C^{\mathrm{T}} \otimes A)\mathrm{vec}(B), \\ \mathrm{trace}(AC^{\mathrm{T}}BC) &= [\mathrm{vec}(C)]^{\mathrm{T}}(A \otimes B)[\mathrm{vec}(C)]. \end{aligned}$$

而且, 若 A 是 $k \times k$ 矩阵, B 是 $l \times l$ 矩阵, 它们的特征值分别为 $\alpha_1, \cdots, \alpha_k$ 与 $\beta_1, \cdots, \beta_l$, 则 $A \otimes B$ 具有特征值 $\alpha_i\beta_j$, $i = 1, \cdots, k$, $j = 1, \cdots, l$.

设空间 $\mathbb{R}^n$ 与 $\mathcal{S}^p$ 分别赋予数积 $x \cdot y := x^{\mathrm{T}}y$ 与 $A \bullet B := \mathrm{trace}(AB)$. 注意到, $\mathcal{S}_-^p$ 关于上述数积的极锥是锥 $\mathcal{S}_+^p := \{A \in \mathcal{S}^p : A \succeq 0\}$, 它是正半定矩阵构成的锥.

5.3.1 负半定矩阵锥的几何

这一节讨论锥 $\mathcal{S}_-^p$ 的某些几何性质, 因为 $\mathcal{S}_+^p = -\mathcal{S}_-^p$, 正半定矩阵锥 $\mathcal{S}_+^p$ 的性质是类似的, 锥 $\mathcal{S}_-^p$ 是闭的凸的, 且可以表示为下述形式:

$$\mathcal{S}_-^p = \{Y \in \mathcal{S}^p : \lambda_{\max}(Y) \leqslant 0\},$$

其中 $\lambda_{\max}(Y)$ 记矩阵 Y 的最大特征值, 因此可用函数 $\lambda_{\max} : \mathcal{S}^p \to \mathbb{R}$ 的微分结构计算 $\mathcal{S}_-^p$ 的切集.

函数 $\lambda_{\max}(\cdot)$ 是凸的, 其一阶与二阶方向导数的计算可参见例 4.20 与例 4.145. 考虑矩阵 $A \in \mathcal{S}^p$. 令 $\alpha := \lambda_{\max}(A)$, $E := [e_1, \cdots, e_s]$ 是 $p \times s$ 矩阵, 它的列 $e_1, \cdots, e_s$ 构成了对应于 A 的最大特征值的特征向量空间的一组直交基, 即 $AE = \alpha E$, $E^{\mathrm{T}}E = I_s$, s 是 A 的最大特征值的重数. 则

$$\lambda'_{\max}(A, H) = \lambda_{\max}(E^{\mathrm{T}}HE). \tag{5.159}$$

所以, $\lambda_{\max}(\cdot)$ 在 A 处是可微的充分必要条件是它的最大特征值的重数等于 1, 此种情况 $\nabla\lambda_{\max}(A)=e_1e_1^{\mathrm{T}}$. 进一步, 令 k 是 $s\times s$ 矩阵 $E^{\mathrm{T}}HE$ 的最大特征值的重数, F 是 $s\times k$ 矩阵, 它的列 $f_1,\cdots,f_k$ 构成了对应 $E^{\mathrm{T}}HE$ 的最大特征值的特征向量空间的一组直交基. 则

$$\lambda''_{\max}(A;H,W)=\lambda_{\max}(F^{\mathrm{T}}E^{\mathrm{T}}[W-2H(A-\alpha I_p)^{\dagger}H]EF). \tag{5.160}$$

设 A 是负半定矩阵且满足 $\lambda_{\max}(A)=0$, 即 A 是集合 $\mathcal{S}_-^p$ 的边界上的点, 因此, 矩阵 E 的列生成 A 的零空间. 由公式 (5.159), $\mathcal{S}_-^p$ 在点 A 处的切锥可以写成下述形式 (见例 2.65):

$$T_{\mathcal{S}_-^p}(A)=\{H\in\mathcal{S}^p:\ E^{\mathrm{T}}HE\preceq 0\}. \tag{5.161}$$

注意到, 若 $\lambda_{\max}(A)<0$, 则 A 属于 $\mathcal{S}_-^p$ 的内部, 此种情况, $T_{\mathcal{S}_-^p}(A)=\mathcal{S}^p$. 这一切锥还可以如下地表示. 设 N 是 $p\times(p-s)$ 矩阵, 它的列由对应于非零特征值的 A 的正交的特征向量构成. 注意到 $A=N\Lambda N^{\mathrm{T}}$, $A^{\dagger}=N\Lambda^{-1}N^{\mathrm{T}}$, 其中 Λ 是 $(p-s)\times(p-s)$ 的对角矩阵, 它的对角元素是 A 的非零的特征值, $p\times p$ 矩阵 $[N,E]$ 是正交的. 由于 $[N,E]$ 是正交矩阵, 任何矩阵 $H\in\mathcal{S}^p$ 均可以表示为

$$\begin{aligned}H&=[N,E]\begin{bmatrix}C_{11}&C_{12}\\C_{21}&C_{22}\end{bmatrix}\begin{bmatrix}N^{\mathrm{T}}\\E^{\mathrm{T}}\end{bmatrix}\\&=NC_{11}N^{\mathrm{T}}+NC_{12}E^{\mathrm{T}}+EC_{21}N^{\mathrm{T}}+EC_{22}E^{\mathrm{T}},\end{aligned} \tag{5.162}$$

其中 $C_{11}:=N^{\mathrm{T}}HN$, $C_{12}:=N^{\mathrm{T}}HE$, $C_{22}:=E^{\mathrm{T}}HE$ 且 $C_{21}=C_{12}^{\mathrm{T}}$. 于是有 $H\in T_{\mathcal{S}_-^p}(A)$ 当且仅当 $C_{22}\preceq 0$.

用上述表示 (5.162), 还可以给出雷达锥$\mathcal{R}_{\mathcal{S}_-^p}(A)$ 的下述刻画. 有

$$A+tH=[N,E]\begin{bmatrix}\Lambda+tC_{11}&tC_{12}\\tC_{21}&tC_{22}\end{bmatrix}\begin{bmatrix}N^{\mathrm{T}}\\E^{\mathrm{T}}\end{bmatrix}. \tag{5.163}$$

因为矩阵 Λ 是负定的, 所以对充分小的 $t>0$, $\Lambda+tC_{11}$ 是负定的. 因此, 对于 $t>0$ 充分小, 矩阵 $A+tH$ 是负半定的充分必要条件是 $(tC_{22})-(tC_{12})^{\mathrm{T}}(\Lambda+tC_{11})^{-1}(tC_{12})$ 是负半定的, 即当且仅当

$$C_{22}-tC_{12}^{\mathrm{T}}(\Lambda+tC_{11})^{-1}C_{12}\preceq 0.$$

因为上述矩阵表示的第二项是 $O(t)$ 阶的, 得到对 $t>0$ 充分小, 这一矩阵是负半定的, 因而 $H\in\mathcal{R}_{\mathcal{S}_-^p}(A)$ 当且仅当 $C_{22}\preceq 0$, 且矩阵 C_{21} 的每一列均属于由矩阵 C_{22} 的列生成的空间.

锥 K 的最大线性子空间被称为其线空间, 记为 $\text{lin}(K)$. 由公式 (5.161), 有

$$\text{lin}(T_{\mathcal{S}_-^p}(A)) = \{H \in \mathcal{S}^p : \ E^{\mathrm{T}}HE = 0\}.$$

等价地, 若 H 由 (5.162) 式给出, 这一线空间由方程 $C_{22} = 0$ 定义. 另一方面, 由雷达锥 $\mathcal{R}_{\mathcal{S}_-^p}(A)$ 的上述描述, 这一雷达锥的线空间可由 $C_{22} = 0$ 且 $C_{12} = 0$ 定义, 或表示为

$$\text{lin}(\mathcal{R}_{\mathcal{S}_-^p}(A)) = \{NBN^{\mathrm{T}} : \ B \in \mathcal{S}^{p-s}\}.$$

考虑集合

$$\mathcal{F} := \{Z \in \mathcal{S}^p : \ Z = NBN^{\mathrm{T}}, \ B \in \mathcal{S}_-^{p-s}\}. \tag{5.164}$$

这一集合即 $\mathcal{S}_-^p$ 与雷达锥$\mathcal{R}_{\mathcal{S}_-^p}(A)$ 的线空间的交集. 集合 $\mathcal{F}$ 构成凸集 $\mathcal{S}_-^p$ 的维数为 $(p-s)(p-s+1)/2$ 的面. 注意到这一集合还可以写为下述形式:

$$\mathcal{F} = \{Z \in \mathcal{S}_-^p : \ Z \bullet (EE^{\mathrm{T}}) = 0\},$$

且矩阵 A 属于 $\mathcal{F}$ 的相对内部.

现在设 $H \in \mathcal{S}^p$ 满足 $\lambda'_{\max}(A, H) = 0$, 因此有 $H \in T_{\mathcal{S}_-^p}(A)$, 相对应的矩阵 F 的列生成 $E^{\mathrm{T}}HE$ 的零空间, 则

$$T^2_{\mathcal{S}_-^p}(A, H) = \{W \in \mathcal{S}^p : \ F^{\mathrm{T}}E^{\mathrm{T}}WEF \preceq 2F^{\mathrm{T}}E^{\mathrm{T}}HA^{\dagger}HEF\} \tag{5.165}$$

(见例 3.40). 注意到, 若 $\lambda'_{\max}(A, H) < 0$, 则对所有 $t > 0$ 充分小, $A + tH$ 是负定的, 因此 $T^2_{\mathcal{S}_-^p}(A, H) = \mathcal{S}^p$.

函数 $\lambda_{\max}(\cdot)$ 是二阶上图正则的, 因此集合 $\mathcal{S}_-^p$ 在任何点 $A \in \mathcal{S}_-^p$ 处均是二阶正则的 (见 3.3.4 节, 尤其是例 3.98).

命题 5.68　令 $A \in \mathcal{S}_-^p$, $H \in T_{\mathcal{S}_-^p}(A)$, 则 $0 \in T^2_{\mathcal{S}_-^p}(A, H)$ 当且仅当 H 属于雷达锥$\mathcal{R}_{\mathcal{S}_-^p}(A)$.

证明　令 $K := \mathcal{S}_-^p$. 显然, 若 $H \in \mathcal{R}_K(A)$, 则 $0 \in T^2_K(A, H)$. 下证相反的结论也是正确的. 若 $\lambda'_{\max}(A, H) < 0$, 则 $T^2_K(A, H) = \mathcal{S}^p$, 因此 $0 \in T^2_K(A, H)$. 设 $\lambda'_{\max}(A, H) = 0$. 根据 (5.165) 得 $0 \in T^2_K(A, H)$ 当且仅当 $F^{\mathrm{T}}E^{\mathrm{T}}HA^{\dagger}HEF \succeq 0$. 因为矩阵 A 是负半定的, 矩阵 $F^{\mathrm{T}}E^{\mathrm{T}}HA^{\dagger}HEF$ 也是负半定的, 因此 $0 \in T^2_K(A, H)$ 当且仅当 $F^{\mathrm{T}}E^{\mathrm{T}}HA^{\dagger}HEF = 0$. 设 N 是 $p \times (p-s)$ 矩阵, 它的列由对应于非零特征值的 A 的直交特征向量构成. 于是 $0 \in T^2_K(A, H)$ 当且仅当 $(N^{\mathrm{T}}HEF)^{\mathrm{T}}\Lambda^{-1}N^{\mathrm{T}}HEF = 0$, 或等价地, 当且仅当 $N^{\mathrm{T}}HEF = 0$. 与 (5.162) 相结合, 得到 $0 \in T^2_K(A, H)$ 当且仅当 $C_{12}F = 0$, 或等价地, $0 = F^{\mathrm{T}}C_{12}^{\mathrm{T}} = F^{\mathrm{T}}C_{21}$, 即 C_{12} 的值空间与 F 的值空间是正交的, 它属于由 C_{22} 的列生成的空间. 由这一命题前面的讨论, 刻画了方向 $H \in \mathcal{R}_K(A)$. 结论得证. □

上述结果意味着对 $A \in \mathcal{S}_-^p$, $H \in \mathcal{S}^p$, 若 $\operatorname{dist}(A+tH,\ \mathcal{S}_-^p) = o(t^2)$, $t > 0$, 则对所有充分小的 $t > 0$, 有 $A + tH \in \mathcal{S}_-^p$.

对 $r \in \{0, \cdots, p\}$, 用 $\mathcal{W}_r$ 记由 $\mathcal{S}^p$ 中的所有秩为 r 的矩阵构成的子集. 集合 $\mathcal{W}_r$ 在下述定义下是局部闭的: 对 $\mathcal{W}_r$ 中任何 A, 存在 $\varepsilon > 0$ 满足 $\mathcal{W}_r \cap \bar{B}(A,\varepsilon)$ 是闭的($\bar{B}(A,\varepsilon)$ 记以 A 为中心, 半径为 ε 的闭球). 还有, 集合 $\mathcal{W}_r$ 构成了线性空间 $\mathcal{S}^p$ 的光滑流形.

命题 5.69 集合 $\mathcal{W}_r$ 是光滑流形, $\mathcal{W}_r$ 在点 $A \in \mathcal{W}_r$ 处的切空间可以写成如下的形式:

$$T_{\mathcal{W}_r}(A) = \{Z \in \mathcal{S}^p : e_i^{\mathrm{T}} Z e_j = 0,\ 1 \leqslant i \leqslant j \leqslant p - r\}, \tag{5.166}$$

其中 $e_1, \cdots, e_{p-r}$ 是矩阵 A 的零空间的一组基.

证明 令 $A \in \mathcal{W}_r$, $E := [e_1, \cdots, e_{p-r}]$ 是由 A 的零空间的一组正交基构成的 $p \times (p-r)$ 矩阵. 考虑在例 3.98 中构造的映射 $\Xi : \mathcal{S}^p \to \mathcal{S}^{p-r}$, 它具有下述性质:

(i) 这一映射在 A 的邻域内是光滑的 (无穷次可微的, 甚至是解析的).

(ii) $\Xi(A) = E^{\mathrm{T}} A E$, 因此 $\Xi(A)$ 是空间 $\mathcal{S}^{p-r}$ 的零矩阵.

(iii) $D\Xi(A)H = E^{\mathrm{T}} H E$, 因此 $D\Xi(A) : \mathcal{S}^p \to \mathcal{S}^{p-r}$ 是映上的.

由这一映射的构造还可以得到, 存在 A 的邻域 $\mathcal{V}_A$ 满足 $\mathcal{W}_r \cap \mathcal{V}_A = \Xi^{-1}(0) \cap \mathcal{V}_A$. 因为 $D\Xi(A)$ 是映上的, 由隐函数定理, $\mathcal{W}_r$ 是 A 的邻域内的光滑流形, 且

$$T_{\mathcal{W}_r}(A) = \{Z \in \mathcal{S}^p : D\Xi(A)Z = 0\}.$$

由关于 $D\Xi(A)$ 的上述公式, 得到

$$T_{\mathcal{W}_r}(A) = \{Z \in \mathcal{S}^p : E^{\mathrm{T}} Z E = 0\}, \tag{5.167}$$

这等价于 (5.166). □

注意, 由 (5.166) 可得

$$\dim \mathcal{W}_r = p(p+1)/2 - (p-r)(p-r+1)/2, \tag{5.168}$$

且若 $A \in \mathcal{S}_-^p$, 则 $T_{\mathcal{W}_r}(A)$ 与切锥$T_{\mathcal{S}_-^p}(A)$ 的线空间是重合的.

下面的横截性的概念借自微分几何.

定义 5.70 称光滑 (连续可微) 映射 $G : \mathbb{R}^n \to \mathcal{S}^p$ 与光滑流形 $\mathcal{W}_r$ 横截地相交于点 $x \in \mathbb{R}^n$, 记为 $G \overline{\pitchfork}_x \mathcal{W}_r$, 若或者 $G(x) \notin \mathcal{W}_r$ 或者

$$T_{\mathcal{W}_r}(G(x)) + DG(x)\mathbb{R}^n = \mathcal{S}^p. \tag{5.169}$$

若对所有的 $x \in \mathbb{R}^n$, 均有 $G \overline{\pitchfork}_x \mathcal{W}_r$, 则称 G 与 $\mathcal{W}_r$ 横截地相交, 记为 $G \overline{\pitchfork} \mathcal{W}_r$.

横截性在小的扰动下是稳定的，即若 $G\overline{\pitchfork}\mathcal{W}_r$, $G':\mathbb{R}^n\to\mathcal{S}^p$ 是在 C^1 范数下充分接近于 G 的另外一个光滑映射，则 $G'\overline{\pitchfork}\mathcal{W}_r$. 尤其，若 $G(x):=A_0+\sum_{i=1}^n x_iA_i$, $G'(x):=A_0'+\sum_{i=1}^n x_iA_i'$ 是两个仿射映射且 $G\overline{\pitchfork}\mathcal{W}_r$, 则当 A_i' 与 $A_i(i=1,\cdots,n)$, 充分接近时，有 $G'\overline{\pitchfork}\mathcal{W}_r$.

在下述意义下，横截性是通有性质. 设 U 是有限维的向量空间，对映射 $\mathcal{G}:\mathbb{R}^n\times U\to\mathcal{S}^p$, 考虑映射 $G_u(\cdot):=\mathcal{G}(\cdot,u)$, $u\in U$. 令 G 是这一参数类映射，即 $G(\cdot):=\mathcal{G}(\cdot,\bar{u})$, 其中 $\bar{u}\in U$ 是某一元素. 设映射 $\mathcal{G}$ 是 C^∞ 光滑的，即 $\mathcal{G}(x,u)$ 在 $\mathbb{R}^n\times U$ 上是无穷次可微的，映射 G① 与 $\mathcal{W}_r$ 是横截地相交的，则对几乎每一 $u\in U$, 映射 G_u 均与 $\mathcal{W}_r$ 横截相交，即满足 G_u 不与 $\mathcal{W}_r$ 横截相交的 $u\in U$ 构成了空间 U 的 Lebesgue 测度为零的集合.

命题 5.71　令 $G(x)\in\mathcal{W}_r$, 即 $\operatorname{rank}G(x)=r$, 用 $e_1,\cdots,e_{p-r}$ 记矩阵 $G(x)$ 的零空间的一组基，令 $G_i:=\partial G(x)/\partial x_i$, 则 $G\overline{\pitchfork}_x\mathcal{W}_r$ 的充分必要条件是下述的 n 维向量

$$v_{ij}:=(e_i^{\mathrm{T}}G_1e_j,\cdots,e_i^{\mathrm{T}}G_ne_j),\quad 1\leqslant i\leqslant j\leqslant p-r$$

是线性无关的.

证明　在方程 (5.169) 的两边取直交补，得到 $G\overline{\pitchfork}_x\mathcal{W}_r$ 当且仅当下述条件成立

$$[T_{\mathcal{W}_r}(A)]^\perp\cap[DG(x)\mathbb{R}^n]^\perp=\{0\},\tag{5.170}$$

其中 $A:=G(x)$. 由公式 (5.166), 空间 $T_{\mathcal{W}_r}(A)$ 的直交补由向量 $e_ie_j^{\mathrm{T}}+e_je_i^{\mathrm{T}}$, $1\leqslant i\leqslant j\leqslant p-r$ 生成. 因为 $DG(x)h=\sum_{k=1}^n h_kG_k$, 有

$$[DG(x)\mathbb{R}^n]^\perp=\{W\in\mathcal{S}^p:\ W\bullet G_k=0,\ k=1,\cdots,n\}.$$

因此，条件 (5.170) 成立当且仅当下述以 α_{ij}, $1\leqslant i\leqslant j\leqslant p-r$ 为未知数的方程组

$$\sum_{1\leqslant i\leqslant j\leqslant p-r}\alpha_{ij}(e_ie_j^{\mathrm{T}}+e_je_i^{\mathrm{T}})\bullet G_k=0,\quad k=1,\cdots,n$$

只有零解. 这等价于向量 ν_{ij}, $1\leqslant i\leqslant j\leqslant p-r$ 的线性无关性.　□

由 (5.169), 若 $G(x)\in\mathcal{W}_r$ 且 $G\overline{\pitchfork}_x\mathcal{W}_r$, 则下述维数条件：

$$\dim\mathcal{W}_r+n\geqslant\dim\mathcal{W}_r+\dim(DG(x)\mathbb{R}^n)\geqslant\dim\mathcal{S}^p.$$

成立. 所以，通常地 (在上述解释的含义下), 由 (5.168), 对任何点 $x\in\mathbb{R}^n$, $r:=\operatorname{rank}(G(x))$, 下述不等式

① 原著中为 $\mathcal{G}$.

$$(p-r)(p-r+1)/2 \leqslant n \tag{5.171}$$

成立, 即令 $r \in \{0,\cdots,p\}$ 使得 (5.171) 不成立. 设 $G(\cdot)$ 是参数类 $\mathcal{G}(\cdot,u)$ 之一, 其中 $\mathcal{G}$ 是 C^∞ 光滑的且 $\mathcal{G}\overline{\pitchfork}\mathcal{W}_r$, 则对几乎每一 $u \in U$, 对所有的 $x \in \mathbb{R}^n$, 均有 $G_u(x) \notin \mathcal{W}_r$.

例如, 考虑仿射映射 $G(x) := A_0 + \sum_{i=1}^n x_i A_i$. 可视这一映射为以矩阵 A_0 为参数的一类仿射映射, 相应的映射 $\mathcal{G}$ 定义为 $\mathcal{G}(x,A) := A + \sum_{i=1}^n x_i A_i$, 其中矩阵 $A_1,\cdots,A_n$ 是固定的. 因为微分 $D\mathcal{G}(x,A)$ 映 $\mathbb{R}^n \times \mathcal{S}^p$ 到 $\mathcal{S}^p$ 上, 所以在任何点 (x,A) 处, $\mathcal{G}$ 与 $\mathcal{W}_r$ 均是横截地相交的, 即 $\mathcal{G}\overline{\pitchfork}\mathcal{W}_r$. 结果得到, 若 $r \in \{0,\cdots,p\}$ 满足不等式 (5.171) 不真, 则对几乎每一 $A \in \mathcal{S}^p$, 对所有的 x, 矩阵 $A + \sum_{i=1}^n x_i A_i$ 的秩均不同于 r.

称 SDP 问题 (P) 的可行点 x 是非退化的, 若 $G\overline{\pitchfork}_x\mathcal{W}_r$, 其中 $r := \operatorname{rank}G(x)$. 非退化性的这一定义与例 4.79 中给出的是一样的. 由上述讨论知道, 非退化性是稳定的通有性质.

现在考虑线性的 SDP 问题 (5.157). 设矩阵 $A_1,\cdots,A_n$ 是线性无关的. (5.157) 的可行域是凸的, 其目标函数是线性的. 设 $\bar{x}$ 是 (5.157) 的可行点, 即 $G(\bar{x}) \in \mathcal{S}^p_-$ 且 $r := \operatorname{rank}(G(\bar{x}))$, 有矩阵 $A = G(\bar{x})$ 属于由 (5.164) 给出的 $\mathcal{S}^p_-$ 相对应的面 $\mathcal{F}$ 的相对内部. 因此, 如果 $\mathcal{F}$ 的维数与由 G 的值空间给出的仿射空间的维数之和大于 $\mathcal{S}^p$ 的维数, 则 $\bar{x}$ 不会是 (5.157) 的可行集的极点. 因为 $\mathcal{F}$ 的维数是 $r(1+r)/2$, 有, 若 $\bar{x}$ 是 (5.157) 可行集的极点, 则

$$r(1+r)/2 \leqslant p(p+1)/2 - n. \tag{5.172}$$

所以对于线性的 SDP 问题, 若 $\bar{x}$ 是其最优解集的极点, 则也是其可行集的极点, 上述不等式 (5.172) 成立. 尤其, 若 $n \geqslant p+1$ 且问题 (5.157) 有非空的有界的最优解集, 则它有最优解 $\bar{x}$ 满足 $\operatorname{rank}G(\bar{x}) \leqslant p-2$.

5.3.2 矩阵凸性

在空间 $\mathcal{S}^p$ 中关于锥 $\mathcal{S}^p_+$ 的偏序被称为 Löwner 偏序, 即对 $A, B \in \mathcal{S}^p$, $A \succeq B$ 当且仅当 $A - B$ 是正半定矩阵. 称映射 $G(x)$ (在凸集 Q 上) 是矩阵凸的, 若它关于 Löwner 偏序是凸的. 这意味着, 对任何 $t \in [0,1]$ 及任意 $x_1, x_2 \in \mathbb{R}^n$(任意的 $x_1, x_2 \in Q$), 不等式

$$tG(x_1) + (1-t)G(x_2) \succeq G(tx_1 + (1-t)x_2) \tag{5.173}$$

成立. 这是 2.3.5 节中讨论的凸映射的特殊情况. SDP 问题 (P) 称为是凸的若 $f(x)$ 是凸的, 集合 Q 是凸的, 映射 $G(x)$ 是矩阵凸的.

命题 5.72 对于映射 $G: \mathbb{R}^n \to \mathcal{S}^p$, 下述结论成立:

(i) $G(x)$ 是矩阵凸的当且仅当对任何 $z \in \mathbb{R}^p$, 实值函数 $\psi(x) := z^{\mathrm{T}}G(x)z$ 是凸的.

(ii) 如果 $G(x)$ 的每一元素 $g_{ij}(x)$, $i,j=1,\cdots,p$ 是二次连续可微函数, 则 $G(x)$ 是矩阵凸的充分必要条件是对任何 $(z_1,\cdots,z_p)\in\mathbb{R}^p$ 及 $x\in\mathbb{R}^n$, $n\times n$ 矩阵 $\sum_{i,j=1}^{p} z_i z_j \nabla^2 g_{ij}(x)$ 是正半定矩阵.

(iii) 若 $G(x)$ 是二次连续可微的, 且 $np\times np$ 块矩阵 $H(x) := [\nabla^2 g_{ij}(x)]_{i,j=1}^{p}$ 对任何 $x\in\mathbb{R}^n$ 均是正半定的, 则 $G(x)$ 是矩阵凸的.

(iv) 若 $G(x)$ 是矩阵凸的, 则它的最大特征值函数 $\phi(x) := \lambda_{\max}(G(x))$ 是凸的.

证明 条件 (5.173) 等价于

$$tz^{\mathrm{T}}G(x_1)z + (1-t)z^{\mathrm{T}}G(x_2)z \geqslant z^{\mathrm{T}}G(tx_1+(1-t)x_2)z, \quad \forall x\in\mathbb{R}^p. \tag{5.174}$$

上述条件意味着对任何 $z\in\mathbb{R}^n$, 函数 $z^{\mathrm{T}}G(\cdot)z$ 是凸的, 证得结论 (i).

二次连续可微实值函数 $\psi(x)$ 是凸的充要条件是对任何 x, Hesse 阵 $\nabla^2\psi(x)$ 是正半定, 又由 (i), 结论 (ii) 得证.

对向量 $z\in\mathbb{R}^p$ 与 $a\in\mathbb{R}^n$, 考虑 $pn\times 1$ 向量 $z\otimes a$, 有

$$(z\otimes a)^{\mathrm{T}}H(x)(z\otimes a) = \sum_{i,j=1}^{p} z_i z_j a^{\mathrm{T}}[\nabla^2 g_{ij}(x)]a. \tag{5.175}$$

若 $H(x)$ 是半正定的, 则等式 (5.175) 的左端是非负的. 这意味着, 对任何 $(z_1,\cdots,z_p)$, 显然 $\sum_{i,j=1}^{p} z_i z_j \nabla^2 g_{ij}(x)$ 是正半定的, 因此由 (ii) 可得到 (iii).

有

$$\lambda_{\max}(G(x)) = \sup_{\|z\|=1} z^{\mathrm{T}}G(x)z.$$

因此, 由 (i), $\lambda_{\max}(G(\cdot))$ 是凸函数的极大值函数, 因此是凸的. 证得 (iv). □

注 5.73 根据 (5.175), 由上述证明可得, 若 $G(x)$ 是二次连续可微的, 则它是矩阵凸的当且仅当对任何 $z\in\mathbb{R}^n$ 与 $a\in\mathbb{R}^p$, $(z\otimes a)^{\mathrm{T}}H(x)(z\otimes a)\geqslant 0$. 等价地, 可把这一条件写为 $M\bullet H(x)\geqslant 0$ 对任何 $M\in\mathcal{C}$ 均成立, 其中 $\mathcal{C}$ 是由形式为 $(zz^{\mathrm{T}})\otimes(aa^{\mathrm{T}})$ 的矩阵构成的 $pn\times pn$ 对称矩阵的闭凸锥. 所以, 这一条件等价于条件 $H(x)\in -\mathcal{C}^-$. 因为锥 $\mathcal{C}$ 被严格地包含在锥 $\mathcal{S}_+^{pn}$ 中, 得到 $H(x)$ 的正半定性是 $G(x)$ 的矩阵凸性的充分条件, 但不是必要的.

显然, 任何仿射映射 $G(x) := A_0 + \sum_{i=1}^{n} x_i A_i$ 是矩阵凸的. 下面给出矩阵凸映射的一些其他例子.

例 5.74 考虑二次映射

$$G(x) := A_0 + \sum_{i=1}^{n} x_i A_i + \sum_{i,j=1}^{n} x_i x_j B_{ij}, \tag{5.176}$$

其中 $A_i, B_{ij} \in \mathcal{S}^p$ 是给定的矩阵. 这一映射是矩阵凸的当且仅当对任何 $x \in \mathbb{R}^n, z \in \mathbb{R}^p$, $\sum_{i,j=1}^{n} x_i x_j z^{\mathrm{T}} B_{ij} z \geqslant 0$. 这一条件的充分条件是 $np \times np$ 块矩阵 $[B_{ij}]_{i,j=1}^{n}$ 是正半定的. 像注记 5.73 那样, 可以证明, 这一块矩阵的正半定性对应二次映射 $G(x)$ 的矩阵凸性的充分条件, 但不是必要条件.

每一函数 $g : \mathbb{R} \to \mathbb{R}$ 通过如下构造对应映射 $G : \mathcal{S}^p \to \mathcal{S}^p$. 考虑矩阵 $Z \in \mathcal{S}^p$, 令 $Z = UDU^{\mathrm{T}}$ 是它的谱分解, 其中 U 是正交阵 (即 $UU^{\mathrm{T}} = I_p$), D 是对角矩阵. 可以定义 $G(Z) := UG(D)U^{\mathrm{T}}$, 其中 $G(D)$ 是对角阵, 对角元素是 $g(d_{11}), \cdots, g(d_{pp})$. 注意到, 若 $g(\cdot)$ 是多项式, 如 $g(t) := a_0 + a_1 t + \cdots + a_m t^m$, 则上述定义的映射 $G(\cdot)$ 由相应的矩阵多项式 $G(Z) = a_0 I_p + a_1 Z + \cdots + a_m Z^m$ 给出. 一个自然的问题是, 是否 $g(\cdot)$ 是凸的, 则其相应的映射 $G(\cdot)$ 是矩阵凸的. 很遗憾, 通常来讲这是不真的. 下面给出这样的矩阵凸映射的两个简单的例子.

例 5.75 考虑映射 $G(Z) = Z^2$. 这一映射是矩阵凸的. 由命题 5.72(i), 为了证明凸性, 只需证明, 对任何 $A, B \in \mathcal{S}^p$, 对任何 $z \in \mathbb{R}^p$, 实值函数 $\psi(t) := z^{\mathrm{T}} G(A+tB) z$ 是凸的, 有 $\psi(t) = z^{\mathrm{T}} A^2 z + 2t(z^{\mathrm{T}} ABz) + t^2(z^{\mathrm{T}} B^2 z)$. 因为矩阵 B^2 总是正半定的, 有 $z^{\mathrm{T}} B^2 z \geqslant 0$, 因此 $\psi(t)$ 事实上是凸函数.

例 5.76 映射 $G(Z) := Z^{-1}$ 在正定矩阵集合 $\mathcal{S}^p_{++}$ 上是矩阵凸的. 事实上, 对 $A \in \mathcal{S}^p_{++}, B \in \mathcal{S}^p, z \in \mathbb{R}^p$, 考虑实值函数 $\psi(t) := z^{\mathrm{T}} G(A+tB) z$, 有二阶导数 $\psi''(0)$ 等于 $2z^{\mathrm{T}}(A^{-1}BA^{-1}BA^{-1})z$. 因为 A^{-1} 是正定的且 $(BA^{-1}z)^{\mathrm{T}} = z^{\mathrm{T}} A^{-1} B$, 有 $\psi''(0) \geqslant 0$. 因为对任意正定矩阵 A, 这一结论都是成立的, 只要 $A + tB$ 停留于 $\mathcal{S}^p_{++}$ 中, $\psi(t)$ 就是凸的. 这证得 Z^{-1} 在集合 $\mathcal{S}^p_{++}$ 上是矩阵凸的.

5.3.3 对偶性

问题 (5.156) 的 Lagrange 函数可以写为下述形式:

$$L(x, \Omega) := f(x) + \Omega \bullet G(x), \quad (x, \Omega) \in \mathbb{R}^n \times \mathcal{S}^p.$$

因此,(5.156) 的 (Lagrange) 对偶问题是

$$\text{(D)} \qquad \max_{\Omega \succeq 0} \{ \inf_{x \in Q} L(x, \Omega) \}. \tag{5.177}$$

若目标函数是线性的, 约束映射是仿射的, 则对偶问题可以显式地表示出来. 例

如, 在线性半定规划问题 (5.157) 的情况, 有

$$\inf_{x\in\mathbb{R}^n} L(x,\Omega)=\begin{cases}\Omega\bullet A_0, & \text{若 } c_i+\Omega\bullet A_i=0, i=1,\cdots,n,\\ -\infty, & \text{否则.}\end{cases}$$

所以, (5.157) 的对偶是

$$\max_{\Omega\succeq 0}\Omega\bullet A_0 \quad \text{s. t.} \quad \Omega\bullet A_i+c_i=0, \quad i=1,\cdots,n. \tag{5.178}$$

注意到, 我们可认为上述问题为锥约束问题的特殊情形, 其中 $f(\Omega):=\Omega\bullet A_0$, $Q:=\mathcal{S}^p_+$, $K:=\{0\}\subset\mathbb{R}^n$, $G(\Omega):=(\Omega\bullet A_1+c_1,\cdots,\Omega\bullet A_n+c_n)$. 它的对偶问题与原始问题 (5.157) 重合. 因此, 对偶对问题 (5.157) 与 (5.158) 间存在完全的对称性, 哪一个称为原始问题, 哪一个称为对偶问题是随意的.

令 $\overline{\Omega}$ 是问题 (5.178) 的可行点, 秩 $\rho:=\operatorname{rank}\overline{\Omega}$. 由 5.3.1 节的讨论, 点 $\overline{\Omega}$ 是非退化的当且仅当由方程 $\Omega\bullet A_i+c_i=0, i=1,\cdots,n$ 定义的仿射空间与 $\mathcal{W}_\rho$ 横截地相交. 即

$$T_{\mathcal{W}_\rho}(\overline{\Omega})+\{\Omega:\Omega\bullet A_i=0, i=1,\cdots,n\}=\mathcal{S}^p, \tag{5.179}$$

或等价地,

$$[T_{\mathcal{W}_\rho}(\overline{\Omega})]^\perp\cap\operatorname{Sp}(A_1,\cdots,A_n)=\{0\}, \tag{5.180}$$

其中 Sp $(A_1,\cdots,A_n)$ 记由矩阵 $A_1,\cdots,A_n$ 张成的线性空间. 令 $\eta_1,\cdots,\eta_{p-\rho}$ 是生成矩阵 $\overline{\Omega}$ 的零空间的直交向量, 则由公式 (5.166), 与 $T_{\mathcal{W}_\rho}(\overline{\Omega})$ 直交的空间由矩阵 $H_{ij}:=\eta_i\eta_j^{\mathrm{T}}+\eta_j\eta_i^{\mathrm{T}}, 1\leqslant i\leqslant j\leqslant p-\rho$ 生成. 进一步, 设矩阵 $A_1,\cdots,A_n$ 是线性无关的, 则点 $\overline{\Omega}$ 是非退化的当且仅当矩阵 H_{ij}, $A_k, 1\leqslant i\leqslant j\leqslant p-\rho, 1\leqslant k\leqslant n$ 是线性无关的, 只有当

$$(p-\rho)(p-\rho+1)/2+n\leqslant p(p+1)/2 \tag{5.181}$$

成立时, 这才可能成立. 所以, 上述不等式是 $\overline{\Omega}$ 的非退化性的必要性条件.

$(\bar{x},\overline{\Omega})$ 被称为 Lagrange 函数 $L(x,\Omega)$ 在集合 $Q\times\mathcal{S}^p_+$ 上的鞍点, 若

$$\bar{x}\in\operatorname*{argmin}_{x\in Q}L(x,\overline{\Omega}), \quad \overline{\Omega}\in\operatorname*{argmax}_{\Omega\in\mathcal{S}^p_+}L(\bar{x},\Omega). \tag{5.182}$$

(5.182) 的第二条件意味着 $G(\bar{x})\in\mathcal{S}^p_-, \overline{\Omega}\in\mathcal{S}^p_+$, 互补条件 $\overline{\Omega}\bullet G(\bar{x})=0$ 成立. 于是得到条件 (5.182) 等价于

$$\bar{x}\in\operatorname*{argmin}_{x\in Q}L(x,\overline{\Omega}), \quad \overline{\Omega}\bullet G(\bar{x})=0, \quad G(\bar{x})\preceq 0, \quad \overline{\Omega}\succeq 0. \tag{5.183}$$

由 (Lagrange) 对偶的一般理论, 有下述结论.

命题 5.77 令 (P) 与 (D) 分别是原始问题 (5.156) 与对偶问题(5.177), 则 val(D)$\leqslant$val(P). 进一步, val(P)=val(D), 且 $\bar{x}$ 与 $\overline{\Omega}$ 分别是 (P) 与 (D) 的最优解的充分必要条件是 $(\bar{x},\overline{\Omega})$ 为 $L(x,\Omega)$ 在 $Q\times S_+^p$ 上的鞍点, 即当且仅当条件 (5.183) 成立.

由 (5.183) 的后三个条件可推出 $\text{rank}G(\bar{x})+\text{rank}\overline{\Omega}\leqslant p$. 称满足最优性条件 (5.183) 的 $\bar{x}$ 与 $\overline{\Omega}$ 是严格互补的, 若

$$\text{rank}G(\bar{x})+\text{rank}\overline{\Omega}=p. \tag{5.184}$$

在线性规划的情况, 假如原始问题或对偶问题的可行集是非空的, 原始问题与对偶问题间没有对偶间隙. 还有, 如果其最优值还是有限的, 则线性规划问题总是具有最优解的. 下面例子表明, 对于线性半定规划问题, 这些性质不总是成立的.

例 5.78 考虑线性半定规划问题:

$$\min x_1 \quad \text{s. t.} \quad \begin{bmatrix} -x_1 & 1 \\ 1 & -x_2 \end{bmatrix}\preceq 0. \tag{5.185}$$

这一问题的可行集是 $\{(x_1,x_2):x_1\geqslant 0,x_2\geqslant 0,x_1x_2\geqslant 1\}$, 因此, 它的最优值是 0 且此问题没有最优解. 这一问题的对偶问题是

$$\max\ 2w_{12} \quad \text{s. t.} \quad \begin{bmatrix} 1 & w_{12} \\ w_{21} & 1 \end{bmatrix}\succeq 0. \tag{5.186}$$

它的可行集包含一点, 满足 $w_{12}=w_{21}=0$, 它也是其最优解, 因此其最优值是 0. 在这个例子中, 原始问题与对偶问题间没有对偶间隙, 尽管原始问题没有最优解.

例 5.79 考虑线性半定规划问题:

$$\min\ -x_2 \quad \text{s. t.} \quad \begin{bmatrix} x_2-a & 0 & 0 \\ 0 & x_1 & x_2 \\ 0 & x_2 & 0 \end{bmatrix}\preceq 0, \tag{5.187}$$

其中 $a>0$ 是给定的数. 这一问题的对偶问题是

$$\max\ -aw_{11} \quad \text{s. t.} \quad \Omega\succeq 0, \quad w_{22}=0, \quad w_{11}+2w_{23}=1. \tag{5.188}$$

原始问题的可行集是 $\{(x_1,x_2):x_1\leqslant 0,x_2=0\}$, 因此其最优值是 0. 另一方面, 对偶问题的任何可行解 Ω 满足 $w_{11}=1$, 因此, 对偶问题的最优值是 $-a$. 所以, 在这个例子中, 对偶间隙是 a.

在线性半定规划问题 (5.157) 与其对偶 (5.178) 的情况下, $(\bar{x},\overline{\Omega})$ 是鞍点的充分必要条件是 $\bar{x}$ 与 $\overline{\Omega}$ 分别是 (5.157) 与 (5.178) 的可行解, 且互补条件

$$\overline{\Omega}\Big(A_0+\sum_{i=1}^n\bar{x}_iA_i\Big)=0 \tag{5.189}$$

成立. 上述条件对应着 (5.183) 的互补条件, $\overline{\Omega}$ 的可行性对应着 (5.183) 的第一与第四条件, $\bar{x}$ 的可行性是 (5.183) 的第三条件. 注意到, 因为 $\overline{\Omega}\succeq 0$ 且 $A_0+\sum_{i=1}^{n}\bar{x}_iA_i\preceq 0$, 互补性条件 (5.189) 等价于 $\overline{\Omega}\cdot\left(A_0+\sum_{i=1}^{n}\bar{x}_iA_i\right)=0$.

现在从共轭对偶性(见 2.5.3 节) 的观点来讨论原始与对偶问题. 与原始问题 (5.156) 相联系的 (标准) 参数化问题是

$$\min_{x\in Q} f(x) \quad \text{s. t.} \quad G(x)+Y\preceq 0, \tag{5.190}$$

它依赖于参数矩阵 $Y\in\mathcal{S}^p$. 将这一问题记为 (P_Y), 用 $v(Y)$ 记它的最优值, 即 $v(Y):=\mathrm{val}(\mathrm{P}_Y)$. 问题 (P_0) 与原始问题 (P) 重合且 val(P)$=v(0)$. 在目前的情况, 函数 $v(\cdot)$ 的共轭由下式

$$-v^*(Y^*)=\begin{cases}\inf\limits_{x\in Q} L(x,Y^*), & 若Y^*\succeq 0,\\ -\infty, & 否则\end{cases} \tag{5.191}$$

给出. 因此, 对偶问题(5.177) 可以表示为在约束 $\Omega\succeq 0$ 之下极大化 $-v^*(\Omega)$ 问题, 进一步, val(P)$=v(0)$ 且 val(D)$=v^{**}(0)$. 由共轭对偶性的一般性理论 (见命题 2.141 与定理 2.142), 下述结果成立.

命题 5.80 下述性质成立:

(i) 若 val(D) 是有限的, 则 $\mathcal{S}(\mathrm{D})=\partial v^{**}(0)$.

(ii) 若 $v(Y)$ 在 $Y=0$ 处是次可微的, 则原始问题与对偶问题间没有对偶间隙, 且 $\mathcal{S}(\mathrm{D})=\partial v(0)$.

(iii) 若 val(P)=val(D) 是有限的, 则 (可能空的) 集合 $\mathcal{S}(\mathrm{D})$, 对偶问题的最优解集, 与 $\partial v(0)$ 重合.

现在考虑约束规范

$$0\in\mathrm{int}\{G(Q)-\mathcal{S}^p_-\}, \tag{5.192}$$

它意味着对 $0\in\mathcal{S}^p$ 的邻域中所有的 Y, $v(Y)<+\infty$ 成立. 若问题 (P) 是凸的, 则最优值函数 $v(\cdot)$ 是凸的. 还注意到, 集合 $\mathcal{S}^p_-$ 的内部由负定矩阵的全体构成, 因此问题 (P) 的 (广义)Slater 条件(定义 2.105) 意味着存在一点 $\bar{x}\in Q$ 满足 $G(\bar{x})\prec 0$. 因为 $\mathcal{S}^p_-$ 的内部非空, 所以若映射 $G(x)$ 是矩阵凸的, 则 Slater 条件等价于约束规范(5.192)(见命题 2.106).

由共轭对偶性理论得到下述结果 (见定理 2.165 与命题 2.166).

定理 5.81 设原始问题 (P) 是凸的. 若问题 (P) 的 Slater 条件成立, 则原始与对偶问题间没有对偶间隙, 进一步, 若它们的公共最优值是有限的, 则对偶问题的最优解集合 $\mathcal{S}(\mathrm{D})$ 是非空的且有界的.

相反地, 若对偶问题具有非空有界的最优解集, 则问题 (P) 的 Slater 条件成立且原始与对偶问题间不存在对偶间隙.

若问题 (P) 是凸的, Slater 条件成立且 val(P) 是有限的, 则 $v(Y)$ 在 $Y=0$ 处是连续的次可微的, 且 $\partial v(0)=\mathcal{S}(\mathrm{D})$. 因此, 下述结果成立.

命题 5.82 设问题 (P) 是凸的, Slater 条件成立, 且 val(P) 是有限的, 则 $v(Y)$ 在 $Y=0$ 处是连续的方向可微的, 且对任何 $Z\in\mathcal{S}^p$,

$$v'(0,Z)=\sup_{\Omega\in\mathcal{S}(\mathrm{D})}\Omega\bullet Z. \tag{5.193}$$

因为若 (P) 是凸的, $v(\cdot)$ 也是凸的, 因而是局部 Lipschitz 连续的, 得到上述命题的最优值函数 $v(Y)$ 在 $Y=0$ 处实际上是 Fréchet 意义下方向可微的.

5.3.4 一阶最优性条件

这一节讨论 SDP 问题 (5.156) 的一阶最优性条件. 设目标函数 $f(x)$ 与约束映射 $G(x)$ 是连续可微的. 为简单起见, 还设集合 Q 与整个空间 $\mathbb{R}^n$ 重合.

设问题 (P) 是凸的, 考虑最优性条件 (5.183). 因为 $L(\cdot,\overline{\Omega})$ 是凸的, 所以 $\bar{x}$ 是 $L(\cdot,\overline{\Omega})$ 的极小点的充分必要条件是 $\nabla_x L(\bar{x},\overline{\Omega})=0$. 因此, 这些条件可以表示为下述形式:

$$\nabla_x L(\bar{x},\overline{\Omega})=0,\quad \overline{\Omega}\bullet G(\bar{x})=0,\quad G(\bar{x})\preceq 0,\quad \overline{\Omega}\succeq 0. \tag{5.194}$$

对线性 SDP 问题 (5.157), 上述条件可简化为作为原始问题 (5.157) 与对偶问题 (5.178) 的点的 $A_0+\sum_{i=1}^{n}\bar{x}_iA_i$ 与 $\overline{\Omega}$ 的可行性及互补性条件 (5.189). 用 $\Lambda(\bar{x})$ 记满足条件 (5.194) 的所有 Lagrange 乘子矩阵 $\overline{\Omega}$ 的集合. 在凸情形下, 若 $\Lambda(\bar{x})$ 是非空的, 则它与对偶问题的最优解集合是重合的. 下述结果是定理 5.81 的一个直接的结论.

定理 5.83 设 SDP 问题 (P) 是凸的, $\bar{x}$ 是它的最优解. 若 (P) 的 Slater 条件成立, 则 Lagrange 乘子矩阵的集合 $\Lambda(\bar{x})$ 对 (P) 的任何最优解均是非空有界的且相同的. 相反地, 若 $\Lambda(\bar{x})$ 是非空的且有界的, 则 Slater 条件成立.

因为锥 $\mathcal{S}_-^p$ 具有非空的内部, 由负定矩阵构成, SDP 问题 (P) 的 Robinson 约束规范可以表示为下述形式 (见引理 2.99):

$$\exists h\in\mathbb{R}^n: G(\bar{x})+DG(\bar{x})h\prec 0. \tag{5.195}$$

若映射 $G(x)$ 是矩阵凸的, 则在任何可行点 $\bar{x}$ 处上述条件 (5.195) 均成立的充分必要条件是问题 (P) 的 Slater 条件成立. 条件 (5.195) 表示 SDP 问题 (P) 在 $\bar{x}$ 处的线性化问题的 Slater 条件. 它还可视为推广的 Mangasarian-Fromovitz 约束规范.

由定理 3.9 与命题 3.17 可得到 SDP 问题 (不必要是凸的) 的下述结果.

定理 5.84 设 $\bar{x}$ 是 SDP 问题 (P) 的局部最优解, 则 Lagrange 乘子矩阵集合 $\Lambda(\bar{x})$ 非空且有界的充分必要条件是约束规范(5.195) 成立.

约束规范(5.195) 可以等价地表示为下述形式 (见命题 2.97):

$$DG(\bar{x})\mathbb{R}^n + T_{\mathcal{S}_-^p}(G(\bar{x})) = \mathcal{S}^p. \tag{5.196}$$

设 r 是矩阵 $G(\bar{x})$ 的秩, 即 $G(\bar{x})$ 属于秩为 r 的矩阵的集合 $\mathcal{W}_r$. 因为 $T_{\mathcal{W}_r}(G(\bar{x}))$ 是 $T_{\mathcal{S}_-^p}(G(\bar{x}))$ 的子集, 所以 (5.196) 可由条件:

$$DG(\bar{x})\mathbb{R}^n + T_{\mathcal{W}_r}(G(\bar{x})) = \mathcal{S}^p \tag{5.197}$$

推出. 上述条件 (5.197) 意味着 G 在点 $\bar{x}$ 处与 $\mathcal{W}_r$ 横截地相交, 或换言之, 点 $\bar{x}$ 是非退化的 (见 5.3.1 节). 因此, 若 $\bar{x}$ 是非退化的, 则约束规范 (5.195) 在 $\bar{x}$ 处成立.

称严格互补条件在可行点 $\bar{x}$ 处成立, 若存在 Lagrange 乘子矩阵 $\Omega \in \Lambda(\bar{x})$ 满足

$$\text{rank}(G(\bar{x})) + \text{rank}\Omega = p. \tag{5.198}$$

若 SDP 问题是凸的, 严格互补性的上述定义与 5.3.3 节给出的定义相同. 在当前的情况下, 它也与抽象定义 4.74 等价. 由命题 4.75, 有下述结果.

定理 5.85 设 $\bar{x}$ 是 SDP 问题 (P) 的局部最优解, 则下述结论成立:

(i) 若点 $\bar{x}$ 是非退化的, 则 $\Lambda(\bar{x})$ 是单点集, 即 Lagarange 乘子矩阵存在且唯一.

(ii) 相反地, 若 $\Lambda(\bar{x})$ 是单点集且严格互补条件在 $\bar{x}$ 处成立, 则点 $\bar{x}$ 是非退化的.

若 SDP 问题 (P) 是凸的, 则在假设 $\Lambda(\bar{x})$ 非空的前提下有 $\Lambda(\bar{x})$ 与 (P) 的对偶问题(D) 的最优解集是重合的. 因此, 在凸的情况下, 上述定理还给出了对偶问题 (D) 最优解的唯一性的条件, 即在凸情形下, 若原始问题 (P) 有非退化的最优解, 则对偶问题 (D) 具有唯一的最优解, 若进一步有 (P) 的最优解 $\bar{x}$ 处严格互补条件成立, 则 (D) 具有唯一最优解的充分必要条件是 $\bar{x}$ 为非退化的.

在没有严格互补条件成立的前提下, $\bar{x}$ 的非退化性只是 Lagrange 乘子矩阵唯一性的充分条件, 但却不是必要条件. 现在讨论 Lagrange 乘子矩阵唯一性这一问题. 令 $\Omega \in \Lambda(\bar{x})$, $r := \text{rank}G(\bar{x})$, $\rho := \text{rank}\Omega$, 令 E 是 $p \times (p-r)$ 矩阵, 它的列是正交的且生成矩阵 $G(\bar{x})$ 的零空间, 则由于 $\Omega \succeq 0$ 且互补条件成立, 有 $\Omega = E \circledH E^{\mathrm{T}}$, 其中 $\circledH \in \mathcal{S}_+^{p-r}$. 注意到, $\circledH$ 的秩等于 Ω 的秩 ρ. 回顾严格约束规范的定义 4.46, 将此约束规范应用到线性化约束 $DG(\bar{x})h \in T_{\mathcal{S}_-^p}(G(\bar{x}))$, 可以表示为下述形式:

$$DG(\bar{x})\mathbb{R}^n + T_{\mathcal{S}_-^p}(G(\bar{x})) \cap \Omega^{\perp} = \mathcal{S}^p, \tag{5.199}$$

其中 $\Omega^{\perp}$ 表示其值域与 Ω 的值域垂直的对称矩阵的集合. 上述条件的抽象形式由 (4.125) 给出. 由命题 4.47(还可见命题 4.50) 有, 上述条件是 Lagrange 乘子矩阵 Ω 唯一性的充分条件, 但如例 4.54 所示, 这一条件并不是必要的.

命题 5.86 条件 (5.199) 是 Lagrange 乘子矩阵 Ω 唯一性的充分条件.

由 $\mathcal{S}_-^p$ 的切锥的表达式 (5.161), 有

$$T_{\mathcal{S}_-^p}(G(\bar{x}))\cap\Omega^{\perp}=\{Z\in\mathcal{S}^p: E^{\mathrm{T}}ZE\preceq 0,\quad \textcircled{H}\cdot E^{\mathrm{T}}ZE=0\}. \tag{5.200}$$

因此, 若矩阵 Ⓗ 是正定的, 即严格互补条件成立, 则 (5.200) 右端由方程 $E^{\mathrm{T}}ZE=0$ 给出. 此种情形, $T_{\mathcal{S}_-^p}(G(\bar{x}))\cap\Omega^{\perp}$ 与 $T_{\mathcal{S}_-^p}(G(\bar{x}))$ 的线空间重合. 因此, 在严格互补条件成立的情况下, (5.199) 与横截性条件$G\,\overline{\pitchfork}_{\bar{x}}\,\mathcal{W}_r$ 等价, 或换言之, 等价于点 $\bar{x}$ 的非退化性.

现在设 Ω 的秩 ρ 小于 $p-r$, 即严格互补条件不成立. 可取矩阵 E 满足 $\textcircled{H}=\begin{bmatrix}\textcircled{H}_{11} & 0\\ 0 & 0.\end{bmatrix}$, 其中 $\textcircled{H}_{11}$ 是 $\rho\times\rho$ 的正定矩阵. 令 $E=[E_1,E_2]$ 是矩阵 E 相应的分划, 因此 $\Omega=E_1\textcircled{H}_{11}E_1^{\mathrm{T}}$. 得到

$$T_{\mathcal{S}_-^p}(G(\bar{x}))\cap\Omega^{\perp}=\{Z\in\mathcal{S}^p: E_1^{\mathrm{T}}ZE_1=0, E_1^{\mathrm{T}}ZE_2=0, E_2^{\mathrm{T}}ZE_2\preceq 0\}. \tag{5.201}$$

考虑向量

$$v_{ij}:=(e_i^{\mathrm{T}}G_1e_j,\cdots,e_i^{\mathrm{T}}G_ne_j),\quad i,j=1,\cdots,p-r,$$

其中 $G_k:=\partial G(\bar{x})/\partial x_k$, 指标集合:

$$\mathcal{I}:=\{(i,j):1\leqslant i\leqslant j\leqslant\rho\}\cup\{(i,j):i=1,\cdots,\rho, j=\rho+1,\cdots,p-r\}.$$

条件 (5.199) 可由下述两个条件推出: (i) 向量 v_{ij}, $(i,j)\in\mathcal{I}$ 是线性无关的, (ii) 存在 $h\in\mathbb{R}^n$ 满足 $h\cdot v_{ij}=0,(i,j)\in\mathcal{I}$, 且

$$\sum_{k=1}^{n}h_kE_2^{\mathrm{T}}G_kE_2\succ 0. \tag{5.202}$$

结果得到, 上述条件 (i) 与 (ii) 是 Lagrange 乘子矩阵 Ω 唯一性的充分性条件. 若 $\rho=p-r$, 即严格互补条件成立, 则上述条件 (i) 与 (ii) 可简化为向量 $v_{ij},i,j=1,\cdots,p-r$ 的线性无关性, 它是点 $\bar{x}$ 的非退化性的充分必要条件.

对 SDP 问题, 还可以不需要任何约束规范, 通过引入广义 Lagrange 乘子将最优性条件写出 (见 3.1.2 节). 考虑 SDP 问题 (P) 的广义 Lagrange 函数:

$$L^g(x,\alpha,\Omega):=\alpha f(x)+\Omega\cdot G(x),\quad (\alpha,\Omega)\in\mathbb{R}\times\mathcal{S}^p, \tag{5.203}$$

相应的最优性条件为

$$\begin{aligned}\nabla_xL^g(x,\alpha,\Omega)=0,\quad \Omega\cdot G(x)=0,\quad G(x)\preceq 0,\\ \Omega\succeq 0,\quad \alpha\geqslant 0,\quad (\alpha,\Omega)\neq(0,0).\end{aligned} \tag{5.204}$$

用 $\Lambda^g(x)$ 记满足上述条件 (5.204) 的广义 Lagrange 乘子(α,Ω) 的集合. 因为约束空间, 即 $\mathcal{S}^p$ 是有限维的, 由命题 3.18 得到下述结果

命题 5.87 *若 $\bar{x}$ 是 SDP 问题 (P) 的局部最优解, 则满足条件 (5.204) 的广义 Lagrange 乘子集合 $\Lambda^g(\bar{x})$ 是非空的.*

5.3.5 二阶最优性条件

这一节讨论 SDP 问题 (5.156) 的二阶最优性条件. 设 $f(x)$ 与 $G(x)$ 是二阶连续可微的且 $Q=\mathbb{R}^n$.

对 $A\in\mathcal{S}_-^p, H\in T_{\mathcal{S}_-^p}(A)$, 计算二阶切集 $\mathcal{T}^2:=T^2_{\mathcal{S}_-^p}(A,H)$ 的支撑函数. 由 (5.165) 得到

$$\sigma(\Omega,\mathcal{T}^2)=\sup\{\Omega\bullet W: F^{\mathrm{T}}E^{\mathrm{T}}WEF\preceq 2F^{\mathrm{T}}E^{\mathrm{T}}HA^{\dagger}HEF\}.$$

注意到这一集合是非空的. 还有, 若 $\Omega\bullet A\neq 0$ 或者 $\Omega\bullet H\neq 0$, 则由命题 3.34, Ω 不属于切锥$T_{T_{\mathcal{S}_-^p}(A)}(H)$ 的极锥, 因此有 $\sigma(\Omega,\mathcal{T}^2)=+\infty$. 所以, 设 $\Omega\in\mathcal{S}_+^p$ 满足 $\Omega\bullet A=0$ 且 $\Omega\bullet H=0$, 因而 Ω 可以写为形式 $\Omega=EF\Psi F^{\mathrm{T}}E^{\mathrm{T}}$, 其中 $\Psi\succeq 0$ 为矩阵. 由于 $\Psi\succeq 0$, 若 $Z_1\preceq Z_2$, 则 $\Psi\bullet Z_1\leqslant\Psi\bullet Z_2$. 从而有

$$\begin{aligned}\sigma(\Omega,\mathcal{T}^2)&=\sup\{\Psi\bullet Z: Z\preceq 2F^{\mathrm{T}}E^{\mathrm{T}}HA^{\dagger}HEF\}\\&=\Psi\bullet(2F^{\mathrm{T}}E^{\mathrm{T}}HA^{\dagger}HEF)=2\Omega\bullet(HA^{\dagger}H).\end{aligned}$$

于是得到

$$\sigma(\Omega,\mathcal{T}^2)=\begin{cases}2\Omega\bullet(HA^{\dagger}H), & 若\Omega\succeq 0,\Omega\bullet A=0,\Omega\bullet H=0,\\ +\infty, & 否则.\end{cases}\tag{5.205}$$

设 $\bar{x}$ 是 SDP 问题 (P) 的局部最优解. 令 $\Lambda^g(\bar{x})$ 是相应的广义 Lagrange 乘子的集合, 令

$$C(\bar{x}):=\{h\in\mathbb{R}^n: DG(\bar{x})h\in T_{\mathcal{S}_-^p}(G(\bar{x})), Df(\bar{x})h\leqslant 0\}\tag{5.206}$$

是临界锥. 若 Lagrange 乘子矩阵的集合 $\Lambda(\bar{x})$ 是非空的, 则对任何 $\Omega\in\Lambda(\bar{x})$, 这一临界锥还可以表示为

$$C(\bar{x})=\{h\in\mathbb{R}^n: DG(\bar{x})h\in T_{\mathcal{S}_-^p}(G(\bar{x})), \Omega\bullet DG(\bar{x})h=0\}.\tag{5.207}$$

令 $h\in C(\bar{x}), (\alpha,\Omega)\in\Lambda^g(\bar{x})$. 由一阶最优性条件 (5.204) 有 $\Omega\bullet G(\bar{x})=0$. 若 $\alpha\neq 0$, 则 $\alpha^{-1}\Omega\in\Lambda(\bar{x})$, 因此 $\Omega\bullet DG(\bar{x})h=0$. 若 $\alpha=0$, 则对任何 $h\in\mathbb{R}^n$, 有 $\Omega\bullet DG(\bar{x})h=0$. 由

$$DG(\bar{x})h=\sum_{i=1}^n h_iG_i(\bar{x}),\quad 其中G_i(\bar{x}):=\partial G(\bar{x})/\partial x_i.$$

用公式 (5.205), 相应的 "sigma 项"可表示为下述形式:

$$\sigma(\Omega, \mathcal{T}^2(h)) = -h^{\mathrm{T}}\mathcal{H}(\bar{x}, \Omega)h, \tag{5.208}$$

其中 $\mathcal{T}^2(h) := T^2_{\mathcal{S}^p_-}(G(\bar{x}), DG(\bar{x})h)$, $\mathcal{H}(\bar{x}, \Omega)$ 是具有典型元素:

$$[\mathcal{H}(\bar{x}, \Omega)]_{ij} := -2\Omega \bullet (G_i(\bar{x})[G(\bar{x})]^{\dagger}G_j(\bar{x})) \tag{5.209}$$

的 $n \times n$ 对称矩阵. 等价地, 矩阵 $\mathcal{H}(\bar{x}, \Omega)$ 可以表示为形式:

$$\mathcal{H}(\bar{x}, \Omega) = -2\left(\frac{\partial G(\bar{x})}{\partial x}\right)^{\mathrm{T}} (\Omega \otimes [G(\bar{x})]^{\dagger}) \left(\frac{\partial G(\bar{x})}{\partial x}\right), \tag{5.210}$$

其中 $\partial G(\bar{x})/\partial x$ 记 $p^2 \times n$ Jacobi 阵:

$$\partial G(\bar{x})/\partial x := [\mathrm{vec}G_1(\bar{x}), \cdots, \mathrm{vec}G_n(\bar{x})].$$

这里的 "sigma 项"$\sigma(\Omega, \mathcal{T}^2(h))$ 是 h 的二次函数. 注意到 3.4.4 节的简约性分析, 这是不奇怪的, 见命题 3.136. 注意到, 因为 $G(\bar{x}) \preceq 0$, $\Omega \succeq 0$, 由 (5.210) 可得矩阵 $\mathcal{H}(\bar{x}, \Omega)$ 是正半定的. 由 3.2.2 节的结果, 尤其是定理 3.50, 现在可给出二阶必要性条件. 注意到锥 $\mathcal{S}^p_-$ 具有非空的内部, 因此, 定理 3.50 中 (不用约束规范的) 二阶必要性条件在这里是可应用的. 集合 $\mathcal{S}^p_-$ 是二阶正则的, 尤其, $\mathcal{S}^p_-$ 的外二阶切集与内二阶切集实际上是重合的.

定理 5.88 *设 $\bar{x}$ 是 SDP 问题 (P) 的局部最优解. 则对任何 $h \in C(\bar{x})$, 存在 $(\alpha, \Omega) \in \Lambda^g(\bar{x})$ 满足*

$$h^T\nabla^2_{xx}L^g(\bar{x}, \alpha, \Omega)h + h^{\mathrm{T}}\mathcal{H}(\bar{x}, \Omega)h \geqslant 0. \tag{5.211}$$

也可以以直接的方式推导上述二阶必要性条件. 我们简短地给出这一方法的轮廓. 考虑最大值函数

$$\phi(x) := \max\{f(x) - f(\bar{x}), \lambda_{\max}(G(x))\}, \tag{5.212}$$

有 $G(x) \preceq 0$ 当且仅当 $\lambda_{\max}(G(x)) \leqslant 0$. 因此, 若 $\bar{x}$ 是 (P) 的局部最优解, 则 $\bar{x}$ 是 $\phi(x)$ 的无约束极小点. 函数 $\phi(x)$ 是一阶与二阶方向可微的. 若 $\bar{x}$ 是 $\phi(x)$ 的无约束极小点, 则 $\phi'(\bar{x}, \cdot) \geqslant 0$, 且 $C(\bar{x}) = \{h : \phi'(\bar{x}, h) = 0\}$. 进一步, 由二阶必要性条件 (见命题 3.105), 对所有的 $h \in C(\bar{x})$, 有 $\phi''(\bar{x}; h, \cdot) \geqslant 0$. 由 $\lambda_{\max}(\cdot)$ 的方向导数的公式与链式法则, 可以计算函数 $\phi(\cdot)$ 的方向导数并可推导出上述最优性条件. 注意到, 若最大值函数 $\phi(\cdot)$ 的二阶增长条件在 $\bar{x}$ 处成立, 则 (5.211) 中的符号 " $\geqslant$ " 可被严格不等式符号 " $>$ " 代替. 然而, 没有额外假设时, SDP 问题 (P) 的二阶增长条件不一定能推出最大值函数 $\phi(\cdot)$ 这样的条件成立.

因为矩阵 $\mathcal{H}(\bar{x},\Omega)$ 是正半定的, 项 $h^{\mathrm{T}}\mathcal{H}(\bar{x},\Omega)h$ 总是非负的. 在某种意义下, 这一项代表了集合 $\mathcal{S}_-^p$ 在点 $G(\bar{x})$ 处曲率的贡献. 若 (5.211) 中的 (α,Ω) 满足 $\alpha\neq 0$, 则 $\alpha^{-1}\Omega\in\Lambda(\bar{x})$, 此种情况, 条件 (5.211) 可代替为条件:

$$h^{\mathrm{T}}\nabla_{xx}^2 L(\bar{x},\Omega)h+h^{\mathrm{T}}\mathcal{H}(\bar{x},\Omega)h\geqslant 0. \tag{5.213}$$

尤其, 若约束规范(5.195) 成立, 则集合 $\Lambda(\bar{x})$ 是非空的有界的, 此种情况, 二阶必要性条件可以表示为形式:

$$\sup_{\Omega\in\Lambda(\bar{x})} h^{\mathrm{T}}(\nabla_{xx}^2 L(\bar{x},\Omega)+\mathcal{H}(\bar{x},\Omega))h\geqslant 0,\quad \forall h\in C(\bar{x}). \tag{5.214}$$

因为集合 $\mathcal{S}_-^p$ 是二阶正则的, 所以上述二阶必要条件在下述意义下是 "无间隙" 条件: 相应的二阶充分性条件可以通过将弱不等式 "$\geqslant 0$" 用严格不等式 ">0" 代替得到 (见定理 3.86). 即可得到下述结果.

定理 5.89　令 $\bar{x}$ 是 SDP 问题 (P) 的可行点, 满足集合 $\Lambda^g(\bar{x})$ 是非空的, 则二阶增长条件在 $\bar{x}$ 处成立若下述条件成立: 对任何 $h\in C(\bar{x})\backslash\{0\}$, 存在 $(\alpha,\Omega)\in\Lambda^g(\bar{x})$ 满足

$$h^{\mathrm{T}}(\nabla_{xx}^2 L^g(\bar{x},\alpha,\Omega)+\mathcal{H}(\bar{x},\Omega))h>0. \tag{5.215}$$

若 $\bar{x}$ 是 SDP 问题 (P) 的局部最优解, 则不管约束规范是否成立, 集合 $\Lambda^g(\bar{x})$ 是非空的. 若约束规范(5.195) 成立, 则上述二阶充分性条件可以表示为形式:

$$\sup_{\Omega\in\Lambda(\bar{x})} h^{\mathrm{T}}(\nabla_{xx}^2 L(\bar{x},\Omega)+\mathcal{H}(\bar{x},\Omega))h>0,\quad \forall h\in C(\bar{x})\backslash\{0\}. \tag{5.216}$$

在约束规范(5.195) 下, 上述二阶条件 (5.216) 是二阶增长条件成立的充分必要条件.

在线性 SDP 问题 (5.157) 的情况下, 相应的 (广义)Lagrange 函数的二阶偏导数是零. 因此, 下述结果是定理 5.89 的一个结论.

推论 5.90　令 $\bar{x}$ 是由 (5.157) 给出的线性 SDP 问题 (P) 的最优解, 设 (P) 的 Slater 条件成立, 则 $\bar{x}$ 处的二阶增长条件成立的充分必要条件是条件:

$$\sup_{\Omega\in\mathcal{S}(D)} h^{\mathrm{T}}\mathcal{H}(\bar{x},\Omega)h>0,\quad \forall h\in C(\bar{x})\backslash\{0\} \tag{5.217}$$

成立, 其中 $\mathcal{S}(D)$ 记线性 SDP 问题 (P) 的对偶问题的最优解集.

考虑临界锥$C(\bar{x})$. 若 $\Lambda(\bar{x})$ 是非空的, 由 (5.207) 与切锥的表示 (5.161), 对任何 $\Omega\in\Lambda(\bar{x})$, 有

$$C(\bar{x})=\left\{h\in\mathbb{R}^n:\sum_{i=1}^n h_iE^{\mathrm{T}}G_i(\bar{x})E\preceq 0,\sum_{i=1}^n h_i\Omega\bullet G_i(\bar{x})=0\right\}. \tag{5.218}$$

令 $\rho := \mathrm{rank}\Omega, E = [E_1, E_2]$ 是矩阵 E 的分划, 满足 $\Omega = E_1 \textcircled{H}_{11} E_1^{\mathrm{T}}$, 其中 $\textcircled{H}_{11}$ 是 $\rho \times \rho$ 的正定矩阵. 若 $\rho = p - r$, 即在严格互补条件成立的情形下, 则 $E = E_1$. 有

$$\sum_{i=1}^{n} h_i E^{\mathrm{T}} G_i(\bar{x}) E = \begin{bmatrix} E_1^{\mathrm{T}} M E_2 & E_1^{\mathrm{T}} M E_2 \\ E_2^{\mathrm{T}} M E_1 & E_2^{\mathrm{T}} M E_2 \end{bmatrix},$$

且 $\sum_{i=1}^{n} h_i \Omega \cdot G_i(\bar{x}) = \textcircled{H}_{11} \cdot E_1^{\mathrm{T}} M E_1$, 其中 $M := \sum_{i=1}^{n} h_i G_i(\bar{x})$. 于是得到 $E_1^{\mathrm{T}} M E_1 = 0, E_1^{\mathrm{T}} M E_2 = 0$ 且 $E_2^{\mathrm{T}} M E_2 \preceq 0$. 结果, 临界锥可以写为下述形式 (与 (5.201) 相比较)

$$C(\bar{x}) = \left\{ h : \sum_{i=1}^{n} h_i E_1^{\mathrm{T}} G_i(\bar{x}) E_1 = 0,\ \sum_{i=1}^{n} h_i E_1^{\mathrm{T}} G_i(\bar{x}) E_2 = 0,\ \sum_{i=1}^{n} h_i E_2^{\mathrm{T}} G_i(\bar{x}) E_2 \preceq 0 \right\}. \tag{5.219}$$

尤其, 若严格互补条件成立, 即存在矩阵 $\Omega \in \Lambda(\bar{x})$ 具有秩 $p - r$, 则

$$C(\bar{x}) = \left\{ h \in \mathbb{R}^n : \sum_{i=1}^{n} h_i E^{\mathrm{T}} G_i(\bar{x}) E = 0 \right\}, \tag{5.220}$$

因而此种情况下, $C(\bar{x})$ 是 $\mathbb{R}^n$ 的线性子空间.

注意到, 可能发生 $C(\bar{x}) = \{0\}$ 的情况. 这一情形发生的充分条件是 (5.219) 右端的线性方程仅有零解. 若严格互补条件成立, $C(\bar{x})$ 可以表示为 (5.220) 的形式, 由线性方程定义, 则这些方程仅有零解是充分与必要条件. 若在极小点 $\bar{x}$ 处的临界锥是 $\{0\}$, 则在 $\bar{x}$ 处一阶增长条件成立.

令 N 是 $p \times r$ 矩阵, 它的列由对应于非零特征值的 $G(\bar{x})$ 的正交特征向量组成, 令 $G(\bar{x}) = NDN^{\mathrm{T}}$ 是 $G(\bar{x})$ 的谱分解, 其中 D 是对角阵, 其对角元素由 $G(\bar{x})$ 的非零特征值组成. 由 (5.209) 得矩阵 $\mathcal{H}(\bar{x}, \Omega)$ 的元素还可以表示为

$$[\mathcal{H}(\bar{x}, \Omega)]_{ij} = -2\textcircled{H}_{11} \cdot (E_1^{\mathrm{T}} G_i(\bar{x}) N D^{-1} N^{\mathrm{T}} G_j(\bar{x}) E_1), \tag{5.221}$$

或等价地,

$$\mathcal{H}(\bar{x}, \Omega) = -2V^{\mathrm{T}}(\textcircled{H}_{11} \otimes D^{-1})V, \tag{5.222}$$

其中 $V := [\mathrm{vec}(N^{\mathrm{T}} G_1(\bar{x}) E_1), \cdots, \mathrm{vec}(N^{\mathrm{T}} G_n(\bar{x}) E_1)]$. 因为矩阵 $\textcircled{H}_{11} \otimes D^{-1}$ 是负定的, 所以矩阵 $\mathcal{H}(\bar{x}, \Omega)$ 是非奇异的, 因而是正定的充分必要条件是矩阵 $E_1^{\mathrm{T}} G_i(\bar{x}) N$, $i = 1, \cdots, n$ 是线性无关的.

定理 5.91 令 $\bar{x}$ 是线性 SDP 问题 (5.157) 的最优解, 设严格互补条件在 $\bar{x}$ 处成立, 则二阶增长条件在 $\bar{x}$ 处成立的充分必要条件是最优解 $\bar{x}$ 是唯一的.

证明 由二阶增长条件的定义, 若二阶增长条件在 $\bar{x}$ 成立, 则 $\bar{x}$ 是 (P) 的局部唯一最优解. 因为线性 SDP 问题是凸的, 所以 $\bar{x}$ 是全局唯一的. 以下证明相反的结果也是正确的.

设 $\bar{x}$ 是唯一的, 严格互补条件成立, 即存在 $\Omega \in \Lambda(\bar{x})$ 满足 $\bar{x}$ 与 Ω 是互补的, 则临界锥$C(\bar{x})$ 由满足方程 $\sum_{i=1}^{n} h_i E^{\mathrm{T}} A_i E = 0$ 的向量 $h \in \mathbb{R}^n$ 构成. 由推论 5.90 的二阶条件与公式 (5.222) 可得, 若对任何 $h \in C(\bar{x})\backslash\{0\}, \sum_{i=1}^{n} h_i E^{\mathrm{T}} A_i N \neq 0$, 则二阶增长条件成立, 即若方程组:

$$\sum_{i=1}^{n} h_i E^{\mathrm{T}} A_i E = 0, \quad \sum_{i=1}^{n} h_i E^{\mathrm{T}} A_i N = 0. \tag{5.223}$$

只有零解, 则二阶增长条件成立. 因为 E 与 N 的列张成空间 $\mathbb{R}^n$, 这等价于矩阵 $\{E^{\mathrm{T}} A_i\}, i = 1, \cdots, n$ 的线性无关性.

相反地, 设存在某非零方向 h, 有 (5.223) 成立. 用分解 (5.162) 与相应的展式 (5.163) 可得, 对充分接近于 0 的所有 $t, G(\bar{x} + th)$ 是正半定的. 但这与线性 SDP 问题最优解 $\bar{x}$ 的唯一性矛盾. □

上述定理的结果也可以表示为如下的形式. 在严格互补条件之下, 线性 SDP 问题的最优解处二阶增长条件成立的充分必要条件是其对偶具有非退化的最优解. 如下述例子表明, 在没有严格互补条件成立的情况下, 对线性 SDP 问题, 即使相应的最优解是唯一的, 二阶增长条件也可能不成立.

例 5.92　考虑线性 SDP 问题:

$$\min_{x \in \mathbb{R}^3} f(x) \quad \text{s. t.} \quad \begin{pmatrix} x_1 & x_2 & 0 & 0 \\ x_2 & 1 & 0 & 0 \\ 0 & 0 & x_2 & x_3 \\ 0 & 0 & x_3 & 1. \end{pmatrix} \succeq 0, \tag{5.224}$$

其中 $f(x) := x_1$. 这一问题的可行集由约束 $x_1 \geqslant x_2^2$, $x_2 \geqslant x_3^2$ 定义, 具有唯一解 $\bar{x} = (0,0,0)$. 考虑路径 $x(t) := (t^4, t^2, t)$, $t \geqslant 0$. 这一路径是可行的, 有 $\|x(t) - \bar{x}\| = O(t), f(x(t)) = t^4$. 因此, 二阶增长条件在点 $\bar{x}$ 处是不成立的.

上述问题的对偶是

$$\max_{\Omega \succeq 0} \ -(\omega_{22} + \omega_{44}) \quad \text{s. t.} \quad \omega_{11} = 1, \quad 2\omega_{12} + \omega_{33} = 0, \quad \omega_{34} = 0. \tag{5.225}$$

它具有唯一的最优解, 由对角阵 $\overline{\Omega} := \mathrm{diag}(1,0,0)$ 给出. $G(\bar{x})$ 的秩是 2, $\overline{\Omega}$ 的秩是 1, 因而此例中严格互补条件是不成立的.

5.3.6　稳定性与灵敏度分析

这一节考虑下述形式

$$(\mathrm{P}_u) \qquad \min_{x \in \mathbb{R}^n} f(x, u) \quad \text{s. t.} \quad G(x, u) \preceq 0 \tag{5.226}$$

的参数化的 SDP 问题. 它依赖于在 Banach 空间 U 中的参数向量 u. 这里 $f:\mathbb{R}^n\times U\to\mathbb{R}$, $G:\mathbb{R}^n\times U\to\mathcal{S}^p$, 对给定的参数向量的值 u_0, 问题 (5.226) 与 "非扰动" 问题 (P) 重合, 即 $f(\cdot,u_0)=f(\cdot)$, $G(\cdot,u_0)=G(\cdot)$. 讨论参数化问题 (5.226) 的最优值参数 $v(u)$ 与最优解 $\hat{x}(u)$ 作为 u 的函数在 u_0 的附近的连续性质和微分性质. 这一节假设函数 $f(x,u)$ 与 $G(x,u)$ 是光滑的 (至少是连续可微的) 且问题 (P) 的最优值是有限的.

因为 SDP 问题是二阶正则的, 这里的分析是第 4 章结果的直接的应用. 所需要的 sigma 项用 5.3.5 节中的方法计算. 若问题 (P) 是凸的, 且其参数化由问题 (5.190) 形式的右端项扰动给出, 则最优值函数的方向导数由命题 5.82 刻画.

考虑 Lagrange 函数

$$L(x,\Omega,u):=f(x,u)+\Omega\bullet G(x,u)$$

以及与 (P_u) 相联系的下确界–紧致性条件(见 4.1 节). 在下确界–紧致性条件之下, 对接近 u_0 的所有 u, (P_u) 的最优解集合是非空的有界的. 在某种意义下, Gol′shtein 定理(定理 4.24) 是命题 5.82 的推广. 在处理当前的问题时, 它可表述为下述形式.

定理 5.93 设 (非扰动) SDP 问题 (P) 是凸的, Slater 与下确界–紧致性条件成立, 则最优值函数 $v(u)$ 在 u_0 处是 Hadamard 方向可微的, 且对任何 $d\in U$,

$$v'(u_0,d)=\inf_{x\in\mathcal{S}(\mathrm{P})}\sup_{\Omega\in\mathcal{S}(\mathrm{D})} d\cdot\nabla_u L(x,\Omega,u_0). \tag{5.227}$$

非凸的情形, 分析就变得非常棘手, (5.227) 的类似公式就可能不成立. 此种情况最优值函数的一阶微分性质与给定的最优化问题的二阶性质相关联 (见 4.8 节). 然而, 由定理 4.26, 可得到下述结果.

定理 5.94 设下确界–紧致性条件成立, 且对任何 $x\in\mathcal{S}(\mathrm{P})$, 存在唯一的 Lagrange 乘子矩阵 $\overline{\Omega}(x)$, 即 $\Lambda(x)=\{\overline{\Omega}(x)\}$, 则最优值函数在 u_0 处是 Hadamard 方向可微的且

$$v'(u_0,d)=\inf_{x\in\mathcal{S}(\mathrm{P})} d\cdot\nabla_u L(x,\overline{\Omega}(x),u_0). \tag{5.228}$$

注意到, Lagrange 乘子矩阵的存在性与唯一性可视为约束规范. 因为此种规范下, $\Lambda(x)$ 显然是非空有界的, 因此这一条件可推出广义 Mangasarian-Fromovitz 约束规范(5.195).

现在讨论问题 (P_u) 的最优解 $\hat{x}(u)$ 的连续性与可微性质. 在下确界–紧致性条件之下, 当 $u\to u_0$ 时, $\hat{x}(u)$ 到非扰动问题的最优解集合 $\mathcal{S}(\mathrm{P})$ 的距离收敛到零. 尤其, 若 $\mathcal{S}(\mathrm{P})=\{\bar{x}\}$ 是单点集, 即 (P) 具有唯一的最优解 $\bar{x}$, 则 $u\to u_0$ 时, $\hat{x}(u)\to\bar{x}$. 然而, $\hat{x}(u)$ 收敛到 $\bar{x}$ 的速度可能较 $O(\|u-u_0\|)$ 要慢, 即当 $u\to u_0$ 时, $\|\hat{x}(u)-\bar{x}\|/\|u-u_0\|$ 可能收敛到 ∞, 即使 $\bar{x}$ 处的二阶增长条件的最强形式是成立的 (见例 4.54).

本节的余下部分, 设问题 (P) 的最优解集是单点集, 即 $\mathcal{S}(\mathrm{P}) = \{\bar{x}\}$, 且 $f(x,u)$ 与 $G(x,u)$ 是二次连续可微的. 对给定的方向 $d \in U$ 与 $t \geqslant 0$, 考虑问题 (P_u) 的线性化

$$(\mathrm{PL}_d) \qquad \begin{array}{ll} \min\limits_{h\in\mathbb{R}^n} & Df(\bar{x},u_0)(h,d), \\ \text{s. t.} & DG(\bar{x},u_0)(h,d) \in T_{\mathcal{S}_-^p}(G(\bar{x},u_0)), \end{array} \tag{5.229}$$

其对偶问题为

$$(\mathrm{DL}_d) \qquad \max_{\Omega\in\Lambda(\bar{x})} d\cdot\nabla_u L(\bar{x},\Omega,u_0). \tag{5.230}$$

在抽象的形式下, 这些问题在 4.3.2 节讨论过.

相对应的方向正则性条件 (见 4.2 节) 可以表示为形式:

$$\exists\varepsilon > 0, \exists h \in \mathbb{R}^n : G(\bar{x},u_0) + DG(\bar{x},u_0)(h,\varepsilon d) \prec 0. \tag{5.231}$$

当广义的 Mangasarian-Fromovitz 约束规范(5.195) 成立时, 对任何方向 d, 上述的方向正则性是成立的.

由命题 4.21, 若方向正则性条件 (5.231) 成立, 则上述问题 (PL_d) 与 (DL_d) 的最优值是相同的, 若进一步有集合 $\Lambda(\bar{x})$ 是非空的, 则这些最优值是有限的, 且问题 (DL_d) 具有非空的有界的最优解集.

考虑下述 (强形式的) 二阶条件:

$$\sup_{\Omega\in\mathcal{S}(\mathrm{DL}_d)} h^{\mathrm{T}}(\nabla^2_{xx}L(\bar{x},\Omega) + \mathcal{H}(\bar{x},\Omega))h > 0, \quad \forall h \in C(\bar{x}) \setminus \{0\}, \tag{5.232}$$

其中矩阵 $\mathcal{H}(\bar{x},\Omega)$ 由 (5.221) 与 (5.222) 定义. 当然, 若 $\mathcal{S}(\mathrm{DL}_d) = \Lambda(\bar{x})$, 则上述条件 (5.232) 与二阶条件 (5.216) 是重合的. 尤其, 这种情况发生在 $\Lambda(\bar{x})$ 是单点集, 或 $G(x,u)$ 不依赖于 u 的情况.

现在可以表述 SDP 问题的基本灵敏性定理.

定理 5.95　*令 $\hat{x}(t) := \hat{x}(u_0 + td)$ 是 $u = u_0 + td$, $t \geqslant 0$ 时参数化问题 (P_u) 的最优解, 满足当 $t \downarrow 0$ 时收敛到 $\bar{x}$. 设方向正则性条件 (5.231) 成立, 强二阶充分性条件 (5.232) 成立, 线性化问题 (PL_d) 的最优解集合 $\mathcal{S}(\mathrm{PL}_d)$ 是非空的. 则*

(i) $\hat{x}(u)$ 在 $\bar{x}$ 处沿方向 d 是 Lipschitz 稳定的, 即

$$\|\hat{x}(t) - \bar{x}\| = O(t), \quad t \geqslant 0. \tag{5.233}$$

(ii) 对 $t \geqslant 0$, 最优值函数沿方向 d 具有下述二阶展式

$$v(u_0 + td) = v(u_0) + t\mathrm{val}(\mathrm{DL}_d) + \frac{1}{2}t^2\mathrm{val}(\mathrm{Q}_d) + o(t^2), \tag{5.234}$$

其中 val(Q_d) 是极小–极大问题:

$$\min_{h\in\mathcal{S}(\mathrm{PL}_d)}\max_{\Omega\in\mathcal{S}(\mathrm{DL}_d)}\{D^2L(\bar{x},\Omega,u_0)((h,d),(h,d))+\mathcal{H}(\bar{x},\Omega)(h,d)\} \tag{5.235}$$

的最优值. 而 $D^2L(\bar{x},\Omega,u_0)((h,d),(h,d))$ 为

$$h^{\mathrm{T}}\nabla^2_{xx}L(\bar{x},\Omega,u_0)h+2h^{\mathrm{T}}\nabla^2_{xu}L(\bar{x},\Omega,u_0)d+d^{\mathrm{T}}\nabla^2_{uu}L(\bar{x},\Omega,u_0)d,$$

且

$$\begin{aligned}
\mathcal{H}(\bar{x},\Omega)(h,d)&:=h^{\mathrm{T}}\mathcal{H}_{xx}(\Omega)h+2h^{\mathrm{T}}\mathcal{H}_{xu}(\Omega)d+d^{\mathrm{T}}\mathcal{H}_{uu}(\Omega)d,\\
\mathcal{H}_{xx}(\Omega)&:=-2\left(\frac{\partial G(\bar{x},u_0)}{\partial x}\right)^{\mathrm{T}}(\Omega\otimes[G(\bar{x},u_0)]^{\dagger})\left(\frac{\partial G(\bar{x},u_0)}{\partial x}\right),\\
\mathcal{H}_{xu}(\Omega)&:=-2\left(\frac{\partial G(\bar{x},u_0)}{\partial x}\right)^{\mathrm{T}}(\Omega\otimes[G(\bar{x},u_0)]^{\dagger})\left(\frac{\partial G(\bar{x},u_0)}{\partial u}\right),\\
\mathcal{H}_{uu}(\Omega)&:=-2\left(\frac{\partial G(\bar{x},u_0)}{\partial u}\right)^{\mathrm{T}}(\Omega\otimes[G(\bar{x},u_0)]^{\dagger})\left(\frac{\partial G(\bar{x},u_0)}{\partial u}\right).
\end{aligned}$$

(iii) $t\downarrow 0$ 时 $(\hat{x}(t)-\bar{x})/t$ 的每一聚点均是极小–极大问题 (Q_d) 的最优解. 尤其, 若 (Q_d) 有唯一的最优解 $\bar{h}$, 则

$$\hat{x}(t)=\bar{x}+t\bar{h}+o(t),\quad t\geqslant 0. \tag{5.236}$$

给出以下注记. 若约束映射 $G(x,u)$ 不依赖于 u, 则线性化问题的最优解集 $\mathcal{S}(\mathrm{PL}_d)$ 与临界锥$C(\bar{x})$ 重合. 此种情形下, 若 $\mathcal{S}(\mathrm{PL}_d)$ 是非空的, 则 $C(\bar{x})$ 构成了 $\mathcal{S}(\mathrm{PL}_d)$ 的回收锥. 上述矩阵 $\mathcal{H}_{xx}(\Omega)$ 恰好是由 (5.221) 与 (5.222) 定义的矩阵 $\mathcal{H}(\bar{x},\Omega)$ 相同.

设 Ω 是 (DL_d) 的最优解. 由 $\Omega\in\Lambda(\bar{x})$, 有 $\Omega G(\bar{x},u_0)=0$. 令 E 是矩阵, 它的列是直交的且生成矩阵 $G(\bar{x},u_0)$ 的零空间, 设 $E=[E_1,E_2]$ 是 E 的分划满足 $\Omega=E_1Ⓗ_{11}E_1^{\mathrm{T}}$, 其中 $Ⓗ_{11}$ 是正定矩阵, 则类似于 (5.219) 得到 (对有限维向量 d)

$$\begin{aligned}
\mathcal{S}(\mathrm{PL}_d):=\Big\{h:&\sum_i h_iE_1^{\mathrm{T}}\frac{\partial G(\bar{x},u_0)}{\partial x_i}E_1+\sum_j d_jE_1^{\mathrm{T}}\frac{\partial G(\bar{x},u_0)}{\partial u_j}E_1=0,\\
&\sum_i h_iE_1^{\mathrm{T}}\frac{\partial G(\bar{x},u_0)}{\partial x_i}E_2+\sum_j d_jE_1^{\mathrm{T}}\frac{\partial G(\bar{x},u_0)}{\partial u_j}E_2=0,\\
&\sum_i h_iE_2^{\mathrm{T}}\frac{\partial G(\bar{x},u_0)}{\partial x_i}E_2+\sum_j d_jE_2^{\mathrm{T}}\frac{\partial G(\bar{x},u_0)}{\partial u_j}E_2\preceq 0\Big\}.
\end{aligned} \tag{5.237}$$

这一节的余下部分考虑严格互补条件与非退化性条件成立的情形. 此时 $\Lambda(\bar{x})=\{\bar{\Omega}\}$ 是单点集, $\mathcal{S}(\mathrm{DL}_d)=\{\bar{\Omega}\}$, 且

$$\mathcal{S}(\mathrm{PL}_d)=\Big\{h:\sum_i h_iE^{\mathrm{T}}\frac{\partial G(\bar{x},u_0)}{\partial x_i}E+\sum_j d_jE^{\mathrm{T}}\frac{\partial G(\bar{x},u_0)}{\partial u_j}E=0\Big\}. \tag{5.238}$$

结果, 相对应的由 (5.235) 定义的问题 (Q_d) 变为线性约束的二次函数的极小化问题, 可以求解为封闭的形式.

在严格互补条件与非退化性条件下, 可以用 4.9.1 节的方法分析 SDP 问题 (P_u) 的灵敏度, 处理等式约束问题. 为此目的, 需要 (局部地) 将约束表达为等式约束. 对充分小的 $\varepsilon>0$, 若 $\hat{x}(u)$ 是 (P_u) 的局部最优解且 $\|\hat{x}(u)-\bar{x}\|+\|u-u_0\|\leqslant\varepsilon$, 由非退化性条件 (此条件在很小的扰动下是稳定的), $\hat{x}(u)$ 与唯一的 Lagrange 乘子 $\overline{\Omega}(u)$ 相联系. 若有必要可减少 ε, 有 $\mathrm{rank}G(\hat{x}(u),u)\geqslant\mathrm{rank}G(\bar{x},u_0)$ 且 $\mathrm{rank}\overline{\Omega}(u)\geqslant\mathrm{rank}\ \overline{\Omega}(u_0)$. 由互补性条件可推出

$$\mathrm{rank}\ G(\hat{x}(u),u)+\mathrm{rank}\ \overline{\Omega}(u)\leqslant n,$$

则可得 $\mathrm{rank}G(\hat{x}(u),u)=\mathrm{rank}G(\bar{x},u_0)$, 即 $G(\hat{x}(u),u)\in\mathcal{W}_r$. 注意到, 对充分接近于 u_0 的 u, 约束 $G(\hat{x}(u),u)\in\mathcal{W}_r$ 可推出 $G(\hat{x}(u),u)\preceq 0$. 在严格互补性与非退化性条件之下, 局部地, 问题 (P_u) 可以等价为下述问题:

$$(\mathrm{P}'_u)\qquad \min_{x\in\mathbb{R}^n} f(x,u)\quad \text{s. t.}\quad G(x,u)\in\mathcal{W}_r. \tag{5.239}$$

通过在 $\mathcal{W}_r$ 中引入局部的坐标系统, 可以将约束 $G(x,u)\in\mathcal{W}_r$ 表示为光滑的方程组系统, 结果可以将隐函数定理用于相应的一阶最优性条件系统. 当然, 可用前面的分析得到同样关于最优值与最优解的方向导数的相应公式. 然而, 隐函数定理能够推导出关于最优解局部性质的更强的结论.

可得到, 若严格互补性条件、非退化性条件与二阶充分性条件成立, $f(x,u)$ 与 $G(x,u)$ 是 k 阶连续可微的, 其中 $k\geqslant 2$, 则在 u_0 的邻域内, $\hat{x}(u)$ 与 $\overline{\Omega}(u)$ 是 $(k-1)$ 阶连续可微的. 还有, 此种情形 $D\hat{x}(u_0)d$ 是下述问题的最优解

$$\begin{aligned}&\min_{h\in\mathbb{R}^n}\quad D^2L(\bar{x},\overline{\Omega},u_0)((h,d),(h,d))+\mathcal{H}(\bar{x},\overline{\Omega})(h,d),\\ &\text{s. t.}\quad \sum_i h_iE^{\mathrm{T}}\frac{\partial G(\bar{x},u_0)}{\partial x_i}E+\sum_j d_jE^{\mathrm{T}}\frac{\partial G(\bar{x},u_0)}{\partial u_j}E=0,\end{aligned} \tag{5.240}$$

且 $D^2v(u_0)(d,d)$ 由上述问题的最优值给出.

5.4 半无限规划

这一节考虑如下形式的最优化问题:

$$(\mathrm{P})\qquad \min_{x\in\mathbb{R}^n} f(x)\quad \text{s. t.}\quad g(x,\omega)\leqslant 0,\quad \forall\omega\in\Omega, \tag{5.241}$$

其中 $f:\mathbb{R}^n\to\mathbb{R}$ 与 $g:\mathbb{R}^n\times\Omega\to\mathbb{R}$ 是实值函数, Ω 是紧致的度量空间. 这一节设函数 $f(x)$ 与 $g(x,\omega)$ 是连续的. 用 $x\cdot y$ 记两个向量 $x,y\in\mathbb{R}^n$ 的标准数积. 若 $f(\cdot)$ 在点 x 处是可微的, 则在 x 处它的微分可表示为形式 $Df(x)h=h\cdot\nabla f(x)$.

若集合 Ω 是有限的, 如 $\Omega=\{\omega_1,\cdots,\omega_p\}$, 则上述问题的可行集由有限个不等式约束 $g_i(x)\leqslant 0, i=1,\cdots,p$ 定义, 其中 $g_i(\cdot):=g(\cdot,\omega_i)$, 因此, 此种情况下, (P) 变为标准的非线性规划问题. 若集合 Ω 是无限的, 称上述问题 (P) 为半无限规划问题. 因为 $X:=\mathbb{R}^n$ 是有限维的且约束的个数是无限的, 考虑的就是半无限规划问题. 如果函数 $f(\cdot)$ 与 $g(\cdot,\omega)$ 是线性的, 如 $f(x):=c\cdot x,\ g(x,\omega):=a(\omega)\cdot x+b(\omega)$, 其中 $c\in\mathbb{R}^n, a:\Omega\to\mathbb{R}^n, b:\Omega\to\mathbb{R}$ 是连续的, 称半无限问题是线性的.

半无限问题 (5.241) 可表示为锥约束问题. 即考虑 Banach 空间 $Y:=C(\Omega)$, 非正值函数构成的锥

$$C_-(\Omega):=\{y\in C(\Omega): y(\omega)\leqslant 0, \forall\omega\in\Omega\},$$

与映射 $G(x)(\cdot):=g(x,\cdot)$, 此映射映点 $x\in\mathbb{R}^n$ 到函数 $g(x,\cdot)\in C(\Omega)$. 显然, 问题 (P) 的可行集 Φ 可以定义为形式 $\Phi=\{x\in\mathbb{R}^n: G(x)\in K\}$, 其中 $K:=C_-(\Omega)$.

映射 $G(x)$ 继承了函数 $g(x,\omega)$ 的许多性质. 尤其有 (见命题 2.174) 下述的一些性质:

(i) 若 $g(x,\omega)$ 是连续的, 则映射 $G:\mathbb{R}^n\to C(\Omega)$ 是连续的.

(ii) 若对每一 $\omega\in\Omega$, 函数 $g(\cdot,\omega)$ 是可微的且 $g(x,\omega)$ 与 $\nabla_x g(x,\omega)$ 是连续的, 则 $G(x)$ 是连续可微的, 且

$$[DG(x)h](\cdot)=h\cdot\nabla_x g(x,\cdot), \tag{5.242}$$

(iii) 映射 $G(x)$ 关于非负值函数的锥 $C_+(\Omega)$ 是凸的充分必要条件是对每一 $\omega\in\Omega$, 函数 $g(\cdot,\omega)$ 是凸函数.

下述几节讨论半无限规划问题的对偶性, 一阶与二阶最优性条件与扰动分析. 考虑关于 Banach 空间$Y:=C(\Omega)$ 与锥 $K:=C_-(\Omega)$ 的某些基本事实. 空间 $C(\Omega)$ 的对偶空间是 Ω 上的有限符号的 Borel 测度的空间, 范数由相应测度的全变差给出, 且对 $y\in C(\Omega)$ 与 $\mu\in C(\Omega)^*$,

$$\langle\mu,y\rangle=\int_\Omega y(\omega)\mathrm{d}\mu(\omega).$$

若 μ 是离散测度, 如 $\mu=\sum_{i=1}^m\lambda_i\delta(\omega_i)$, 其中 $\delta(\omega)$ 记质量 1 在点 $\omega\in\Omega$ 的 (Dirac) 测度, 则 $\|\mu\|=\sum_{i=1}^m|\lambda_i|$ 且 $\langle\mu,y\rangle=\sum_{i=1}^m\lambda_i y(\omega_i)$.

$K := C_-(\Omega)$ 的极锥由非负测度 $\mu \in C(\Omega)^*$ 的集合构成, 即 $\mu \in K^-$ 当且仅当 $\mu(A) \geqslant 0$, 对任何 Borel 集 $A \subset \Omega$ (见例 2.63) 成立. 用 $\mu \succeq 0$ 表示测度 μ 是非负的. 还注意到集合 $C_-(\Omega)$ 具有非空的内部且函数 $y \in C(\Omega)$ 属于集合 $C_-(\Omega)$ 的内部当且仅当 $y(\omega) < 0$ 对所有的 $\omega \in \Omega$ 成立. 这是由于 Ω 是紧致的, 若对所有的 $\omega \in \Omega$, 有 $y(\omega) < 0$, 则存在 $\varepsilon > 0$ 满足对所有的 $\omega \in \Omega$ 成立着 $y(\omega) \leqslant -\varepsilon$.

考虑函数 $y \in C_-(\Omega)$,

$$\Delta(y) := \{\omega \in \Omega : y(\omega) = 0\} \tag{5.243}$$

记 y 的接触点集合, 集合 $K := C_-(\Omega)$ 在 y 处的切锥 $T_K(y)$ 可表示为 (见例 2.63)

$$T_K(y) = \{z \in C(\Omega) : z(\omega) \leqslant 0, \forall \omega \in \Delta(y)\}, \tag{5.244}$$

相应的法锥为

$$N_K(y) = \{\mu \in C(\Omega)^* : \operatorname{supp}(\mu) \subset \Delta(y), \mu \succeq 0\}. \tag{5.245}$$

5.4.1 对偶性

问题 (P) 的 Lagrange 函数

$$L(x,\mu) := f(x) + \langle \mu, G(x) \rangle = f(x) + \int_\Omega g(x,\omega)\mathrm{d}\mu(\omega),$$

其中 $(x,\mu) \in \mathbb{R}^n \times C(\Omega)^*$. 所以, (P) 的对偶问题可表示为形式:

$$\text{(D)} \qquad \max_{\mu \succeq 0} \left\{ \psi(\mu) := \inf_{x \in \mathbb{R}^n} L(x,\mu) \right\}, \quad \mu \in C(\Omega)^*. \tag{5.246}$$

尤其, 若问题 (P) 是线性的, 即具有下述形式

$$\text{(LP)} \qquad \min_{x \in \mathbb{R}^n} \ c \cdot x \quad \text{s. t.} \quad a(\omega) \cdot x + b(\omega) \leqslant 0, \quad \forall \omega \in \Omega, \tag{5.247}$$

其对偶问题可以如下表示为更加显式的形式, 即

$$\psi(\mu) = \begin{cases} \displaystyle\int_\Omega b(\omega)\mathrm{d}\mu(\omega), & \text{若} \ \ c + \displaystyle\int_\Omega a(\omega)\mathrm{d}\mu(\omega) = 0, \\ -\infty, & \text{否则}. \end{cases}$$

所以, 线性半无限问题 (LP) 的对偶问题可以表示为

$$\text{(LD)} \qquad \max_{\mu \succeq 0} \int_\Omega b(\omega)\mathrm{d}\mu(\omega) \quad \text{s. t.} \quad c + \int_\Omega a(\omega)\mathrm{d}\mu(\omega) = 0. \tag{5.248}$$

若 Ω 是有限集合, 则上述问题 (LP) 变为线性规划问题. 由定理 5.44 可得, 除非两个问题均不是相容的, 原始问题与对偶问题间不存在对偶间隙. 当集合 Ω 是无限集时, 情况就变得非常复杂. 对于无限集 Ω, 集合 $C_-(\Omega)$ 的切锥与雷达锥不一定重合, 此种情形, 可以构造形式为 (LP) 的线性规划问题, 满足其最优值比它的对偶问题(LD) 的最优值大 (见命题 2.193). 因此, 即使是对线性的半无限规划问题, 对偶间隙也可能存在 (见例 5.102 与 5.103).

现在设问题 (P) 是凸的, 即函数 $f(\cdot)$ 与 $g(\cdot,\omega), \omega \in \Omega$ 是凸的, 回顾 2.5 节发展的对偶理论的某些结果. 考虑问题 (P) 的如下 (标准的) 参数化:

$$(\mathrm{P}_y) \qquad \min_{x\in\mathbb{R}^n} f(x) \quad \text{s. t.} \quad g(x,\omega)+y(\omega)\leqslant 0, \quad \forall \omega \in \Omega, \tag{5.249}$$

它依赖于 $y \in C(\Omega)$, 相应的最优值函数为 $v(y) := \mathrm{val}(\mathrm{P}_y)$, 则有 $\mathrm{val}(\mathrm{P}) = v(0)$, $\mathrm{val}(\mathrm{D}) = v^{**}(0)$, 且由于 (P) 是凸的, 最优值函数 $v(y)$ 是凸的. 根据 Fenchel-Moreau-Rockafellar 定理(定理 2.113), 有 $\mathrm{val}(\mathrm{D}) = \mathrm{cl}\, v(0)$.

$\mathrm{lsc}\, v(0)$ 称为问题 (P) 的次值, 若该次值小于 $+\infty$, 则称 (P) 是次相容的. 等价地, 如果存在收敛到 0 的序列 $\{y_k\} \subset C(\Omega)$(以 $C(\Omega)$ 的范数拓扑收敛, 即 $k\to\infty$ 时 $y_k(\omega) \to 0$ 关于 $\omega \in \Omega$ 是一致的), 满足序列 $\{v(y_k)\}$ 是上方有界的, 则问题 (P) 是次相容的. 当然, 若 (P) 是相容的, 即其可行集非空, 则它是次相容的. 若 (P) 是次相容的, 则 $\mathrm{cl}\, v(0) = \mathrm{lsc}\, v(0)$, 因而有 $\mathrm{val}(\mathrm{D}) = \mathrm{lsc}\, v(0)$. 因此, 可得下述结果 (见定理 2.144).

定理 5.96 设问题 (P) 是凸的, 则

$$\mathrm{val}(\mathrm{D}) = \mathrm{cl}\, v(0). \tag{5.250}$$

进一步, 若还有 (P) 是次相容的, 则

$$\mathrm{val}(\mathrm{D}) = \mathrm{lsc}\, v(0) = \min\{\mathrm{val}(\mathrm{P}), \liminf_{y\to 0} v(y)\} \tag{5.251}$$

且 $\mathrm{val}(\mathrm{P}) = \mathrm{val}(\mathrm{D})$ 当且仅当

$$\mathrm{val}(\mathrm{P}) \leqslant \liminf_{y\to 0} v(y). \tag{5.252}$$

若 $v(0)$ 是有限的, $v(y)$ 在 $y=0$ 处是次可微的, 则称问题 (P) 是平静的. 因为集合 $K := C_-(\Omega)$ 具有非空的内部, 由命题 2.160 可得 (又见命题 2.164) 凸问题(P) 是平静的充分必要条件是 $\mathrm{val}(\mathrm{P})$ 是有限的, 且

$$\liminf_{t\downarrow 0} \frac{v(td)-v(0)}{t} > -\infty, \quad \forall d \in C(\Omega). \tag{5.253}$$

根据命题 2.147, 若 (P) 是平静的, 则 (P) 与 (D) 间不存在对偶间隙且 (D) 的最优解集是非空的. 相反地, 若 val(P) 是有限的且 (P) 与 (D) 间不存在对偶间隙, 则 (D) 有最优解当且仅当 (P) 是平静的.

由集合 $K := C_-(\Omega)$ 的法锥的公式 (5.245), 最优性条件 (2.306) 可以表述如下:

$$\begin{cases} x_0 \in \underset{x\in\mathbb{R}^n}{\operatorname{argmin}}\, L(x,\mu), \\ g(x_0,\omega) \leqslant 0,\ \omega \in \Omega,\ \operatorname{supp}(\mu) \subset \Delta(x_0),\ \mu \succeq 0, \end{cases} \tag{5.254}$$

其中

$$\Delta(x_0) := \{\omega \in \Omega : g(x_0,\omega) = 0\}$$

是在 x_0 处起作用的约束的集合. 由于 $g(x_0,\cdot)$ 是连续的且非正值的, 且 $\mu \succeq 0$, 条件 $\operatorname{supp}(\mu) \subset \Delta(x_0)$ 表示为互补条件 $\int_\Omega g(x_0,\omega)\mathrm{d}\mu(\omega) = 0$.

由 (5.247) 形式描述的线性半无限问题 (LP) 可视为锥线性问题. 因此, 对 (LP) 问题, 上述最优性条件 (5.254) 取下述形式 (见命题 2.191 的方程 (2.359)):

$$\begin{cases} a(\omega)\cdot x_0 + b(\omega) \leqslant 0,\ \omega \in \Omega, \\ c + \displaystyle\int_\Omega a(\omega)\mathrm{d}\mu(\omega) = 0,\ \mu \succeq 0, \\ \displaystyle\int_\omega [a(\omega)\cdot x_0 + b(\omega)]\mathrm{d}\mu(\omega) = 0. \end{cases} \tag{5.255}$$

(5.255) 的前两个条件表示问题 (LP) 与其对偶问题(LD) 的可行性, 最后一个条件是互补条件.

由定理 2.158, 有如下问题 (P) 与 (D) 的强对偶性的充分与必要条件.

定理 5.97 *若 val(P) = val(D), $x_0 \in \mathbb{R}^n$ 与 $\bar{\mu} \in C(\Omega)^*$ 分别是 (P) 与 (D) 的最优解, 则条件 (5.254) 成立. 相反地, 若存在 $x_0 \in \mathbb{R}^n$ 与 $\bar{\mu} \in C(\Omega)^*$ 满足条件 (5.254), 则 x_0 是 (P) 的最优解, $\bar{\mu}$ 是 (D) 的最优解, (P) 与 (D) 间不存在对偶间隙.*

上述结果给出 "无对偶间隙"与最优解的存在性的必要性与充分性条件. 遗憾的是, 对于具体的情况, 不总是很容易地验证像 (5.252) 或 (5.253) 这样的条件. 称 Slater 条件成立, 若存在 $\bar{x}$ 满足 $G(\bar{x}) \in \operatorname{int}(K)$. 当前情况下, Slater 条件对于 (P) 成立可表示为如下形式:

$$\exists \bar{x} \in \mathbb{R}^n \text{ 满足 } g(\bar{x},\omega) < 0, \quad \forall \omega \in \Omega. \tag{5.256}$$

因为集合 $C_-(\Omega)$ 具有非空的内部, 正则性条件 (3.12) 等价于 Slater 条件 (见命题 2.106).

定理 5.98 *若问题 (P) 是凸的且 Slater 条件 (5.256) 成立, 则 val(P)=val(D), 若进一步, val(P) 是有限的, 则对偶问题(D) 的最优解集合是非空的有界的 (以 $C(\Omega)^*$*

的全变差范数拓扑). 相反地, 若 (P) 是凸的, val(P) 是有限的, 且 (D) 的最优解集是非空的有界的, 则 val(P)=val(D) 且 Slater 条件成立.

证明 上述定理可直接由定理 2.165 得到. 因为集合 $C_-(\Omega)$ 具有非空内部, 相反的结论由命题 2.166 得到. □

注意到, 到目前为止还没有用到空间 $X := \mathbb{R}^n$ 维数的有限性.

现在考虑分别由 (5.247) 与 (5.248) 给出的线性半无限问题 (LP) 及其对偶问题(LD). 若分别赋予空间 $C(\Omega)$ 与 $C(\Omega)^*$ 强拓扑与弱 * 拓扑, 则它们成为成对的局部凸的拓扑向量空间. 此种情况, 问题 (LP) 与 (LD) 间存在完整的对称性, 即 (LD) 的对偶问题是 (LP)(见 2.5.6 节). 考虑 (LD) 的参数化

$$(\mathrm{LD}_z) \qquad \min_{\mu \preceq 0} \int_\Omega b(\omega)\mathrm{d}\mu(\omega) \quad \text{s. t.} \quad c - \int_\Omega a(\omega)\mathrm{d}\mu(\omega) = z \tag{5.257}$$

及与之相联系的最优值函数 $w(z) := \mathrm{val}(\mathrm{LD}_z)$. 注意到, 为使 (LD_z) 变成极小化问题, 将 μ 变换为 $-\mu$, 从而 $w(0) = -\mathrm{val}(\mathrm{LD})$. 函数 $w(z)$ 是凸的, 且由于定义在有限维空间 $\mathbb{R}^n$ 上, 它于 $z = 0$ 处次可微的充分必要条件是

$$\liminf_{t \downarrow 0} \frac{w(td) - w(0)}{t} > -\infty, \quad \forall d \in \mathbb{R}^n. \tag{5.258}$$

因此, 由命题 2.147 得, 若 val(LD) 是有限的, 则 val(LP)=val(LD) 且 (LP) 的最优解集是非空的当且仅当条件 (5.258) 是成立的.

还有, 因为最优值函数 $w(z)$ 定义在有限维空间上, 在 $z = 0$ 处连续的充分必要条件是 $0 \in \mathrm{int}(\mathrm{dom}\, w)$, 有

$$\mathrm{dom}\, w = \left\{ z \in \mathbb{R}^n : z = c + \int_\Omega a(\omega)\mathrm{d}\mu(\omega),\ \mu \succeq 0,\ \mu \in C(\Omega)^* \right\}.$$

因此, 正则性条件 $0 \in \mathrm{int}(\mathrm{dom}\, w)$ 取下述形式:

$$0 \in \mathrm{int}\left\{ z \in \mathbb{R}^n : z = c + \int_\Omega a(\omega)\mathrm{d}\mu(\omega),\ \mu \succeq 0,\ \mu \in C(\Omega)^* \right\}. \tag{5.259}$$

定理 5.99 若条件 (5.259) 成立, 则 val(LP) = val(LD), 若进一步, val(LP) 是有限的, 则 (LP) 的最优解集合是非空的有界的. 相反地, 若 (LP) 的最优解集是非空的有界的, 则条件 (5.259) 成立且 val(LP)=val(LD).

证明 条件 (5.259) 等价于 $0 \in \mathrm{int}(\mathrm{dom}\, w)$, 进而等价于 $w(z)$ 在 $z = 0$ 处的连续性. 上述定理的直接结论由定理 2.151 得到. 相反地, 设 $\mathcal{S}$(LP) 是非空的有界的. 分别赋予空间 $C(\Omega)$ 与 $C(\Omega)^*$ 强拓扑与弱 * 拓扑, 使得这两个空间成为成对的局部凸的拓扑向量空间. 由命题 2.141 有 $\partial w^{**}(0) = -\mathcal{S}(\mathrm{LP})$ (因为 $w(z)$ 是问题 $-$(LD) 的最优值函数, 所以有减号), 从而 $\partial w^{**}(0)$ 是有界的. 因为 $X := \mathbb{R}^n$ 是有限

维的, 有 $w^{**}(z)$ 在 $z=0$ 处是连续的, 因而 $0 \in \text{int}(\text{dom } w)$. 可推得 (5.259), 因此有 $\text{val(LP)} = \text{val(LD)}$. 证毕. □

考虑离散化的问题. 对给定的有限集合 $\{\omega_1, \cdots, \omega_m\} \subset \Omega$, 考虑问题 (P) 的下述离散化:

$$(\text{P}_m) \qquad \min_{x \in \mathbb{R}^n} f(x) \quad \text{s. t.} \quad g(x, \omega_i) \leqslant 0, \quad i = 1, \cdots, m, \tag{5.260}$$

与其对偶问题:

$$(\text{D}_m) \qquad \max_{\lambda \in \mathbb{R}^m_+} \left\{ \inf_{x \in \mathbb{R}^n} \left[f(x) + \sum_{i=1}^{m} \lambda_i g(x, \omega_i) \right] \right\}. \tag{5.261}$$

注意到, 上述问题 (D_m) 可通过在对偶问题(D) 中只对形式为 $\mu := \sum_{i=1}^{m} \lambda_i \delta(\omega_i)$ 的离散度量取极大化得到. 因为 (P_m) 的可行集包含 (P) 的可行集, 有 $\text{val(P)} \geqslant \text{val}(\text{P}_m)$. 由于离散测度构成了 $C(\Omega)^*$ 的一个子集, 有 $\text{val(D)} \geqslant \text{val}(\text{D}_m)$.

下述命题表明, 若 (P) 是凸的且存在 (P) 的离散化, 具有和 (P) 相同的最优值, 则可找到 (P) 的离散化 (P_m), 具有和 (P) 相同的最优值且满足 $m \leqslant n$. 这一结果的证明基于 Helly 定理(定理 2.182), 该定理表述为: 若 $\mathbb{R}^n$ 中的有限个凸集合满足其中每 $n+1$ 个集合均有非空的交集, 则这族集合的交是非空的.

命题 5.100 *设问题 (P) 是凸的相容的, 且存在具有相同最优值的 (P) 的有限的离散化. 则存在离散化 (P_m) 满足 $\text{val(P)} = \text{val}(\text{P}_m)$ 且 $m \leqslant n$.*

证明 因为 (P) 是相容的, 有 $\text{val(P)} < +\infty$. 若 $\text{val(P)} = -\infty$, 则对任何离散化, 均有 $\text{val}(\text{P}_m) = -\infty$, 因而结论成立. 所以可设 val(P) 是有限的. 令 (P_k) 为对应集合 $\{\omega_1, \cdots, \omega_k\}$ 的 (P) 的离散化, 满足 $\text{val}(\text{P}_k) = \text{val(P)}$. 考虑集合 $A_i := \{x \in \mathbb{R}^n : g(x, \omega_i) \leqslant 0\}$, $i = 1, \cdots, k$, $A_0 := \{x \in \mathbb{R}^n : f(x) < \text{val(P)}\}$. 注意到问题 (P_k) 的可行集是集合 A_i, $i = 1, \cdots, k$ 的交集, 有集合 A_i, $i = 0, 1, \cdots, k$ 是凸的且它们的交集是空集合. 结果, 由 Helly 定理, 这族集合中含有至多 $n+1$ 个集合的子族满足这一子族集合的交是空集. 这一子族应该包含 A_0, 因为否则必有 $\bigcap_{i=1}^{k} A_i = \varnothing$, 这与 (P) 的可行集非空的假设是矛盾的. 令 $A_0, A_{i_1}, \cdots, A_{i_m}$ 构成这一子族, 其中 $m \leqslant n$. 考虑对应集合 $\{\omega_{i_1}, \cdots, \omega_{i_m}\}$ 的离散化问题 $\text{val}(\text{P}_m)$. 由上述构造, 得到 $(\text{P}_m) \geqslant \text{val(P)}$, 因此 $\text{val}(\text{P}_m) = \text{val(P)}$. 证毕. □

如例 5.102 所示, 即使是对线性半无限规划问题, 也可能发生这样的情况, 即对相应离散化集合 $\{\omega_1, \cdots, \omega_m\}$ 的任何选择, 离散化问题 (P_m) 的最优值比 val(P) 小.

例 5.101 考虑线性半无限问题:

$$\min_{(x_1, x_2) \in \mathbb{R}^2} x_2 \quad \text{s. t.} \quad -\omega^2 x_1 - x_2 + 2\omega \leqslant 0, \quad \forall \omega \in [0, 1]. \tag{5.262}$$

这一问题的可行集由不等式

$$x_2 \geqslant \begin{cases} -x_1+2, & \text{若 } x_1 \leqslant 1, \\ 1/x_1, & \text{若 } x_1 \geqslant 1 \end{cases}$$

定义. 注意到, 对任何 $\omega \in (0,1]$, 由方程 $-\omega^2 x_1 - x_2 + 2\omega = 0$ 定义的直线在点 $(1/\omega, \omega)$ 处与曲线 $x_1x_2 = 1$ 相切. 问题 (5.262) 的最优值是零, 这一问题不具有最优解. 对任何包含点 $\omega_1 := 0$ 的任何有限集合 $\{\omega_1, \cdots, \omega_m\} \subset [0,1]$, 相应的离散化问题有最优值零, 若 $m=1$, 无限多个最优解形成了直线 $\{(x_1,x_2): x_2 = 0\}$, 若 $m \geqslant 2$, 则存在 $a>0$, 最优解形成了半直线 $\{(x_1,x_2): x_2=0, x_1 \geqslant a\}$. 若离散化集合不包含 0, 则相应的离散化问题的最优值是 $-\infty$.

问题 (5.262) 的对偶问题是

$$\max_{\mu \succeq 0} \int_0^1 2\omega \mathrm{d}\mu(\omega) \quad \text{s. t.} \quad \int_0^1 \omega^2 \mathrm{d}\mu(\omega) = 0, \quad \int_0^1 \mathrm{d}\mu(\omega) = 1. \tag{5.263}$$

这一问题有唯一的可行点 $\mu = \delta(0)$(在点 $\omega = 0$ 处质量为 1 的测度) 也是最优解. 对偶问题的最优值是 0, 因此问题 (5.262) 与 (5.263) 间不存在对偶间隙. 这不会令人吃惊, 因为原始问题 (5.262) 的 Slater 条件成立.

例 5.102 考虑线性半无限问题:

$$\min_{(x_1,x_2)\in\mathbb{R}^2} x_2 \quad \text{s. t.} \quad -\omega^3 x_1 - \omega x_2 + 2\omega^2 \leqslant 0, \quad \forall \omega \in [0,1]. \tag{5.264}$$

上述问题 (5.264) 的可行集与问题 (5.262) 的可行集是相同的, 因此这一问题也有最优值零. 另一方面, 对任何 (5.264) 的离散化, 相应的离散化问题 (P_m) 的最优值是 $-\infty$.

问题 (5.264) 的对偶问题是

$$\max_{\mu \succeq 0} \int_0^1 2\omega^2 \mathrm{d}\mu(\omega) \quad \text{s. t.} \quad \int_0^1 \omega^3 \mathrm{d}\mu(\omega) = 0, \int_0^1 \omega \mathrm{d}\mu(\omega) = 1. \tag{5.265}$$

这一问题的可行集是空集, 因此它的最优值是 $-\infty$. 所以, (5.264) 与其对偶 (5.265) 间的对偶间隙是无穷大. 注意到问题 (5.264) 的 Slater 条件不成立.

例 5.103 考虑线性的半无限问题:

$$\min \quad x_2 \quad \text{s. t.} \quad -\omega^3 x_1 - \omega x_2 + 2\omega^2 \leqslant 0, \quad \forall \omega \in [0,1], \quad -x_2 - 1 \leqslant 0. \tag{5.266}$$

这一问题由 (5.264) 添加约束 $x_2 \geqslant -1$ 得到. 问题 (5.266) 可视为形式 (5.241) 的线性半无限规划问题, 其中取 $\Omega := [0,1] \cup \{2\} \subset \mathbb{R}$, 若 $\omega \in [0,1]$, $g(x,\omega) := -\omega^3 x_1 - \omega x_2 + 2\omega^2$, 且 $g(x,2) := -x_2 - 1$. 问题 (5.264) 与 (5.266) 具有相同的可行

集与相同的最优值 0. 任何包含最后约束的问题 (5.266) 的离散化问题均有最优值 (-1). 问题 (5.266) 的对偶问题是

$$\begin{aligned}\max_{\mu\succeq 0,\lambda\geqslant 0}\quad & \int_0^1 2\omega^2\mathrm{d}\mu(\omega)-\lambda,\\ \text{s. t.}\quad & \int_0^1 \omega^3\mathrm{d}\mu(\omega)=0,\ \int_0^1 \omega\mathrm{d}\mu(\omega)+\lambda=1.\end{aligned}\tag{5.267}$$

这一对偶问题的可行域是 $\{(\mu,\lambda):\mu=\alpha\delta(0),\lambda=1,\alpha\geqslant 0\}$. 因此, 它的最优值是 (-1) 且每一可行点也是最优解. 问题 (5.266) 与 (5.267) 间存在对偶间隙. 对 (5.266) 的任何离散化 (P_m), 均有 $\mathrm{val}(\mathrm{P}_m)<\mathrm{val}(\mathrm{P})$, 尽管对偶问题具有非空的最优解集.

现在给出存在 (P) 的离散化具有相同最优值的条件.

命题 5.104 *设问题 (P) 是凸的, 存在 $x_0\in\mathbb{R}^n$ 与 $\bar\mu\in C(\Omega)$ 满足最优性条件 (5.254), 则存在点 $\omega_i\in\Omega$, $i=1,\cdots,m$, 其中 $m\leqslant n$, 满足对偶问题(D) 具有支撑为集合 $\{\omega_1,\cdots,\omega_m\}$ 的最优解, 且 $\mathrm{val}(\mathrm{P})=\mathrm{val}(\mathrm{P}_m)=\mathrm{val}(\mathrm{D}_m)=\mathrm{val}(\mathrm{D})$, 其中 (P_m) 与 (D_m) 分别是问题 (P) 与 (D) 的相应的离散化.*

证明 设问题 (P) 是凸的, $x_0\in\mathbb{R}^n$, $\bar\mu\in C(\Omega)$ 满足最优性条件 (5.254). 注意到, 由定理 5.97 有, x_0 与 $\bar\mu$ 分别是 (P) 与 (D) 的最优解且 val(P)=val(D) 是有限值. 还由于对 $\mu\succeq 0$, 函数 $L(\cdot,\mu)$ 是凸的, 条件 (5.254) 因而可表示为下述形式:

$$0\in\partial L(x_0,\mu),\quad g(x_0,\omega)\leqslant 0,\quad \omega\in\Omega,\quad \mathrm{supp}(\mu)\subset\Delta(x_0),\quad \mu\succeq 0.\tag{5.268}$$

令 $\mathcal{M}$ 是满足上述最优性条件且 $\mu(\Omega)=\bar\mu(\Omega)$ 的所有测度 $\mu\in C(\Omega)^*$ 的集合. 由于 $\bar\mu\in\mathcal{M}$, 集合 $\mathcal{M}$ 是非空的凸的有界的且弱 * 闭的, 因而是弱 * 紧致的. 结果, 由 Krein-Milman 定理(定理 2.19), 它至少有一极点. 与命题 2.177 的证明完全一样, 可以证明 $\mathcal{M}$ 的任何极点均有至多含 $n+1$ 个点的支撑. 于是存在点 $\omega_i\in\Omega$, $i=1,\cdots,k$, 其中 $k\leqslant n+1$ 与 $\hat\mu\in\mathcal{M}$ 满足 $\mathrm{supp}(\hat\mu)=\{\omega_1,\cdots,\omega_k\}$. 结果, 可表示为 $\hat\mu=\sum_{i=1}^k\lambda_i\delta(\omega_i)$, 且由最优性条件 (5.268), 存在 $q_0\in\partial f(x_0)$ 与 $q_i\in\partial g(x_0,\omega_i)$, $i=1,\cdots,k$, 满足

$$q_0+\sum_{i=1}^k\lambda_i q_i=0,\quad \omega_i\in\Delta(x_0),\quad \lambda_i\geqslant 0,\quad i=1,\cdots,k.\tag{5.269}$$

进一步, 由于向量 q_i 是 n 维的, 可以选取满足 (5.269) 的乘子 λ_i', 使得其中至多有 n 个是非零的. 令 $\omega_i'\in\{\omega_1,\cdots,\omega_k\}$, $i=1,\cdots,m$, $m\leqslant n$ 是对应于非零 λ_i' 的点, $\mu':=\sum_{i=1}^m\lambda_i'\delta(\omega_i)$ 是相应的离散测度, 则有 μ' 满足最优性条件 (5.268), 因而由定理 5.97 可得 μ' 是 (D) 的最优解.

用 (P_m) 与 (D_m) 记问题 (P) 与集合 $\{\omega'_1,\cdots,\omega'_m\}$ 相联系的离散化与该离散化的对偶. 由于 μ' 是 (D) 与 (D_m) 的最优解, 得 $\mathrm{val}(\mathrm{D})=\mathrm{val}(\mathrm{D}_m)$, 因为 $\mathrm{val}(\mathrm{P})=\mathrm{val}(\mathrm{D})$, $\mathrm{val}(\mathrm{P})\geqslant\mathrm{val}(\mathrm{P}_m)$ 且 $\mathrm{val}(\mathrm{P}_m)\geqslant\mathrm{val}(\mathrm{D}_m)$, 得到 $\mathrm{val}(\mathrm{P})=\mathrm{val}(\mathrm{P}_m)=\mathrm{val}(\mathrm{D}_m)=\mathrm{val}(\mathrm{D})$. □

下述结果基于 Helly 定理(定理 2.181) 的下述形式: 若 A_i, $i\in I$ 是 $\mathbb{R}^n$ 中的一族 (可能是无限) 凸紧致集合, 满足这一族集合中每 $n+1$ 个集合均有一非空的交集, 则这族中的所有集合的交集是非空的.

命题 5.105 *设问题 (P) 是凸的相容的且下述条件成立:*

$$\begin{aligned}&\text{对任何}n+1\text{个点}\omega_1,\cdots,\omega_{n+1}\in\Omega,\text{存在一点}\bar{x}\in\mathbb{R}^n\text{满足}\\&g(\bar{x},\omega_i)<0,\ i=1,\cdots,n+1.\end{aligned}\tag{5.270}$$

则存在点 $\omega_1,\cdots,\omega_n\in\Omega$, 满足

$$\mathrm{val}(\mathrm{P})=\mathrm{val}(\mathrm{P}_n)=\mathrm{val}(\mathrm{D}_n)=\mathrm{val}(\mathrm{D}),$$

其中 (P_n) 与 (D_n) 分别是问题 (P) 与 (D) 的对应的离散化.

证明 因为 (P) 是相容的, 有 $\mathrm{val}(\mathrm{P})<+\infty$. 若 $\mathrm{val}(\mathrm{P})=-\infty$, 则结果显然可得. 因此, 可设 val(P) 是有限的. 需要证明, 存在 (P) 的有限的离散化 (P_n) 满足 $\mathrm{val}(\mathrm{P})=\mathrm{val}(\mathrm{P}_n)$. 由假设 (5.270) 知对每一离散化 (P_n), Slater 条件均是成立的, 这可推出 $\mathrm{val}(\mathrm{P}_n)=\mathrm{val}(\mathrm{D}_n)$, 因此有 $\mathrm{val}(\mathrm{P}_n)=\mathrm{val}(\mathrm{D}_n)=\mathrm{val}(\mathrm{D})$.

考虑集合 $\overline{\Omega}:=\Omega\cup\{\omega_0\}$, 它由集合 Ω 添加一点 ω_0 得到, 定义 $g(\cdot,\omega_0):=f(\cdot)-\mathrm{val}(\mathrm{P})$. 如果取 $\{\omega_0\}$ 是 ω_0 的邻域, 集合 $\overline{\Omega}$ 仍然是紧致的. 对 $m\in\mathbb{N}$ 与 $\omega\in\overline{\Omega}$, 定义集合

$$A_m(\omega):=\{x\in\mathbb{R}^n:\|x\|\leqslant m,g(x',\omega)<0,\forall x'\in x+m^{-1}B\},$$

其中 $B:=\{x:\|x\|<1\}$ 是 $\mathbb{R}^n$ 的开单位球. 观察到集合 $A_m(\omega)$ 是闭的. 事实上, 令 $x_k\in A_m(\omega)$ 且 $x_k\to\hat{x}$. 则 $\|\hat{x}\|\leqslant m$. 进一步, 若 $x'\in\hat{x}+m^{-1}B$, 则对充分大的 k, 有 $\|x_k-\hat{x}\|<m^{-1}-\|x'-\hat{x}\|$, 因此 $x'\in x_k+m^{-1}B$. 这得到 $\hat{x}\in A_m(\omega)$, 从而 $A_m(\omega)$ 是闭的. 很清楚, $A_m(\omega)$ 也是有界的, 因此它是紧致的. 由于函数 $f(\cdot)$ 与 $g(\cdot,\omega)$ 是凸的, 集合 $A_m(\omega)$ 是凸的.

因为对每一 $m\in\mathbb{N}$, 每一点 $x\in\bigcap_{\omega\in\Omega}A_m(\omega)$ 是 (P) 的可行点且 $f(x)<\mathrm{val}(\mathrm{P})$, 集合 $\bigcap_{\omega\in\overline{\Omega}}A_m(\omega)$ 是空集. 结果, 由 Helly 定理(定理 2.181), 对每一 $m\in\mathbb{N}$, 存在 $n+1$ 个点 $\omega_1^m,\cdots,\omega_{n+1}^m\in\overline{\Omega}$ 满足集合 $\bigcap\limits_{i=1}^{n+1}A_m(\omega_i^m)$ 是空集. 因为 $\overline{\Omega}$ 是紧致的, 存在序列 $m_k\to\infty$ 满足 $(\omega_1^{m_k},\cdots,\omega_{n+1}^{m_k})$ 收敛到一点 $(\omega_1,\cdots,\omega_{n+1})\in\overline{\Omega}^{n+1}$. 考虑集合

$$\mathcal{A} := \{x : g(x, \omega_i) < 0,\ i = 1, \cdots, n+1\}.$$

函数 $g(x,\omega), \omega \in \overline{\Omega}$ 是连续的, 若 $x \in \mathcal{A}$, 则对充分大的 $m_k, x \in \bigcap_{i=1}^{n+1} A_{m_k}(\omega_i^{m_k})$, 从而可得集合 $\mathcal{A}$ 是空集. 连同假设 (5.270), 这意味着 $\omega_0 \in \{\omega_1, \cdots, \omega_{n+1}\}$. 令 $\omega_1, \cdots, \omega_n$ 是集合 $\{\omega_1, \cdots, \omega_{n+1}\}$ 除去 ω_0 的点, 得到 $\mathcal{A} = \{x : f(x) < \mathrm{val(P)}, g(x, \omega_i) < 0,\ i = 1, \cdots, n\} = \varnothing$, 从而对任何满足约束 $g(x, \omega_i) < 0,\ i = 1, \cdots, n$ 的 $x \in \mathbb{R}^n$, 均有 $f(x) \geqslant \mathrm{val(P)}$. 因为 $g(\cdot, \omega_i)$ 是凸函数且假设 (5.270) 成立, 集合 $\{x : g(x, \omega_i) \leqslant 0,\ i = 1, \cdots, n\}$ 由集合 $\{x : g(x, \omega_i) < 0,\ i = 1, \cdots, n\}$ 的拓扑包给出. 由 $f(\cdot)$ 的连续性, 对任何满足约束 $g(x, \omega_i) \leqslant 0,\ i = 1, \cdots, n$ 的 $x \in \mathbb{R}^n$ 均有 $f(x) \geqslant \mathrm{val(P)}$, 证得结论. □

注意到 Slater 条件 (5.256) 可推出上述命题中用到的条件 (5.270). 由上述命题还可以注意到, 若问题 (P) 是凸的, val(P) 有限且条件 (5.270) 成立, 则 val(P)=val(D) 且对偶问题(D) 具有支撑至多含有 n 个点的最优解, 因而 (P) 是平静的.

如果半无限规划问题是线性的, 即具有 (5.247) 的形式, 则可以得到更多的结果. 观察到线性半无限规划问题(LP) 的任何有限的离散化 (LP_m) 是线性规划问题, 因此, 除非两个问题 (LP_m) 与 (LD_m) 是不相容的, 有 $\mathrm{val(LP}_m) = \mathrm{val(LD}_m)$.

命题 5.106 *设线性半无限规划问题(LP) 的最优值是有限的, 则下述结论是等价的:*

(i) (LP) 有具有相同最优值的有限的离散化.

(ii) 问题 (LP) 是平静的.

(iii) val(LP)=val(LD) 且对偶问题(LD) 具有最优解.

证明 设 (LP) 具有相同最优值的有限离散化 (LP_m). 考虑 (LP) 的标准的参数化:

$$(\mathrm{LP}_y) \qquad \min_{x \in \mathbb{R}^n} \ c \cdot x \quad \text{s. t.} \quad a(\omega) \cdot x + b(\omega) + y(\omega) \leqslant 0, \quad \forall \omega \in \Omega. \tag{5.271}$$

令 (LP_{y^m}) 是 (LP_y) 的相对应的离散化, 有 $\mathrm{val(LP}_y) \geqslant \mathrm{val(LP}_{y^m})$. 由线性规划的理论有线性规划问题 ($\mathrm{LP}_m$) 是平静的. 结果, 由条件 (5.253) 得问题 (LP) 也是平静的. 这证得推出关系 (i)⇒(ii).

推出关系 (ii)⇒(iii) 由一般理论 (见命题 2.147) 得到.

设 val(LP) = val(LD) 且对偶问题 (LD) 有最优解 $\bar{\mu}$, 则可得 (LD) 有具有有限支撑的最优解. 事实上, 考虑所有满足 $c + \int_\Omega a(\omega)\mathrm{d}\mu(\omega) = 0, \int_\Omega b(\omega)\mathrm{d}\mu(\omega) = \int_\Omega b(\omega)\mathrm{d}\bar{\mu}(\omega)$ 且 $\mu(\Omega) = \bar{\mu}(\Omega)$ 的测度 $\mu \succeq 0$. 这是非空的凸的有界的, 因而是弱 * 紧致的集合. 结果, 由 Krein-Milman 定理, 它至少含有一极点, 由构造知这一极点是

(LD) 的最优解. 而且, 因为该集合由 $n+2$ 个线性方程定义, 这样的极点的支撑至多含有 $n+2$ 个点.

令 (LD_m) 与 (LP_m) 是对应于 (LD) 的具有有限支撑的最优解的 (LD) 与 (LP) 的离散化. 有 $\text{val(LP)} \geqslant \text{val}(\text{LP}_m) = \text{val}(\text{LD}_m) = \text{val(LD)}$, 连同假设 val(LP)=val(LD), 得到 $\text{val(LP)} = \text{val}(\text{LP}_m)$. 这证得推出关系 (iii)⇒(i). □

5.4.2 一阶最优性条件

这一节讨论半无限规划问题 (5.241) 在可行点 x_0 处的一阶必要与/或充分性最优性条件. 若在 x_0 处起作用的约束的集合 $\Delta(x_0)$ 是空集, 则 x_0 是 (P) 的可行集的内部点, 此种情形, 为分析 x_0 的附近的局部性质, 问题 (P) 可视为无约束问题, 且标准的 (无约束的) 一阶 (与二阶) 最优性条件可用于点 x_0 处的函数 $f(\cdot)$. 因此, 除非特别声明, 在后续的分析中, 均假设集合 $\Delta(x_0)$ 是非空的.

如果半无限问题 (P) 是凸的, 则 x_0 点处的最优性条件可表示为 (5.254) 的形式, 或等价地, 表示为 (5.268) 形式. 注意到, 因为函数 $f(\cdot)$ 与 $g(\cdot,\omega)$ 设为凸的且连续的, 所以, 若 $\mu = \sum_{i=1}^{m} \lambda_i \delta(\omega_i)$ 是离散的测度, 则

$$\partial L(x,\mu) = \partial f(x) + \sum_{i=1}^{m} \lambda_i \partial g(x,\omega_i), \tag{5.272}$$

其中所有的次微分是关于 x 的.

定理 5.107 设半无限规划问题 (P) 是凸的, x_0 是 (P) 的最优解, Slater 条件 (5.256) 成立, 则

(i) 满足条件 (5.254) 的所有测度 $\mu \in C(\Omega)^*$ 的集合 Λ_0 是非空的凸的且有界的, 而且对 (P) 的任何最优解, 该集合均是相同的.

(ii) 对任何 $m \in \mathbb{N}$, 满足条件 (5.254) 的所有离散测度 $\mu = \sum_{i=1}^{m} \lambda_i \delta(\omega_i)$ 的集合是有界的, 对任何 $m \geqslant n$, 这一集合是非空的, 且对 (P) 的任何最优解, 这一集合均是相同的.

证明 由定理 5.98 得, 在给定的假设下, val(P)=val(D), 对偶问题(D) 的最优解的集合是非空的凸的且有界的. 由定理 5.97, 还得到 Λ_0 与 (D) 的最优解集是重合的. 结论 (i) 得到.

为证明 (ii), 首先观察到, 若 Λ_0 是有界的, 则很显然离散测度的这一子集也是有界的. 进一步, 因为 Λ_0 是弱 * 紧致的, 由 Krein-Milman 定理, 它是它的极点的凸包的 (弱 *) 拓扑包. 所以, 为完成证明只需要证, 若测度 μ 是 Λ_0 的极点, 则它的支撑至多有 n 个点. 最后的结论可按完全与命题 2.177 的证明相同的方式得到证明 (同命题 5.104 的证明相比较). □

现在考虑可微的情形, 即设函数 $f(\cdot)$ 与 $g(\cdot,\omega),\omega\in\Omega$ 是可微的. 若进一步, 这些函数又是凸函数, 即问题 (P) 是凸的, 则 $\partial L(x,\mu)=\{\nabla_x L(x,\mu)\}$, 因此在可行点 x_0 处的最优性条件 (5.254) 可以表示为下述形式:

$$\nabla_x L(x_0,\mu)=0,\quad \text{supp}\,(\mu)\subset\Delta(x_0),\quad \mu\succeq 0, \tag{5.273}$$

其中 $\Delta(x_0)$ 是在 x_0 处起作用的约束的集合. 当问题 (P) 不必要是凸问题时, 上述最优性条件也是有意义的. 尤其, 若 $\mu=\sum_{i=1}^{m}\lambda_i\delta(\omega_i)$ 是离散的测度, 条件 (5.273) 可以表示为下述形式:

$$\nabla f(x_0)+\sum_{i=1}^{m}\lambda_i\nabla_x g(x_0,\omega_i)=0,\quad \omega_i\in\Delta(x_0),\quad \lambda_i\geqslant 0,\quad i=1,\cdots,m. \tag{5.274}$$

在合适的假设下, 可以在积分值内取 Lagrange 函数的导数. 对给定的测度 $\mu\in C(\Omega)^*$, 称函数 $h:\Omega\to\mathbb{R}$ 是 μ 可积的, 若积分 $\int_\Omega h(\omega)\mathrm{d}\mu(\omega)$ 是有定义的, 即 $h(\omega)$ 是 μ 可测的且 $\int_\Omega |h(\omega)|\mathrm{d}\mu(\omega)<+\infty$. 任何连续函数 $h:\Omega\to\mathbb{R}$ 均是 μ 可测的, 因为 Ω 是紧致的, 从而 $h(\omega)$ 是有界的, 这一函数是 μ 可积的.

命题 5.108 设 $x_0\in\mathbb{R}^n,\mu\in C(\Omega)^*$ 是给定的, 考虑函数 $\gamma(x):=\int_\Omega g(x,\omega)\mathrm{d}\mu(\omega)$. 设

(i) 对每一 $x\in\mathbb{R}^n$, 函数 $g(x,\cdot)$ 是 μ 可积的.

(ii) 存在 μ 可积函数 $\kappa(\omega)$ 满足

$$|g(x',\omega)-g(x'',\omega)|\leqslant\kappa(\omega)\|x'-x''\| \tag{5.275}$$

对 x_0 的某一邻域中所有的 x' 与 x'' 以及 $\omega\in\Omega$ 成立.

(iii) 对 μ 几乎处处的 ω, 函数 $g(\cdot,\omega)$ 在 x_0 处是可微的.

则函数 $\gamma(x)$ 在 x_0 的邻域内是 Lipschitz 连续的, 在 x_0 处是可微的, 且

$$\nabla\gamma(x_0)=\int_\Omega\nabla_x g(x_0,\omega)\mathrm{d}\mu(\omega). \tag{5.276}$$

证明 条件 (5.275) 当然意味着 $g(\cdot,\omega)$ 在 x_0 附近是以 $\kappa(\omega)$ 为模 Lipschitz 连续的. 对于在有限维空间上定义的 Lipschitz 连续函数而言, Gâteaux, Fréchet 与 Hadamard 可微性的概念是等价的 (见 2.2.1 节). 由假设 (i) 有函数 $\gamma(x)$ 是有定义的. 由 (ii) 有

$$|\gamma(x')-\gamma(x'')|\leqslant\int_\Omega|g(x',\omega)-g(x'',\omega)|\mathrm{d}\mu(\omega)\leqslant\|x'-x''\|\int_\Omega\kappa(\omega)\mathrm{d}\mu(\omega),$$

因此 $\gamma(x)$ 在 x_0 的邻域内是 Lipschitz 连续的. 考虑 $h \in \mathbb{R}^n$. 由 (ii), 对充分小的 $t>0$ 与 μ 几乎处处的 ω, 有

$$t^{-1}|g(x_0+th,\omega)-g(x_0,\omega)| \leqslant \kappa(\omega)\|h\|.$$

则由 Lebesgue 控制收敛定理有

$$\lim_{t\downarrow 0}\left[\frac{\gamma(x_0+th)-\gamma(x_0)}{t}\right] = \int_\Omega \lim_{t\downarrow 0}\left[\frac{g(x_0+th,\omega)-g(x_0,\omega)}{t}\right]\mathrm{d}\mu(\omega).$$

上述等式的左端是方向导数$\gamma'(x_0,h)$, 所以得到 $\gamma(x)$ 在 x_0 处是方向可微的, 且

$$\gamma'(x_0,h) = \int_\Omega g'_\omega(x_0,h)\mathrm{d}\mu(\omega), \tag{5.277}$$

其中 $g'_\omega(x_0,h)$ 记 $g(\cdot,\omega)$ 在 x_0 处沿方向 h 的方向导数. 由假设 (iii), 有 $g'_\omega(x_0,h) = h\cdot\nabla_x g(x_0,\omega)$ 对 μ 几乎处处的 ω 成立, 因此 $\gamma'(x_0,h)$ 关于 h 是线性的. 结果, $\gamma(x)$ 在 x_0 处是可微的且公式 (5.276) 成立. 再次注意到, 由于 $\gamma(x)$ 在 x_0 附近是 Lipschitz 连续的, $\gamma(x)$ 在 x_0 处是 Fréchet 意义可微的. □

给出注记, 假设 $g(\cdot,\omega)$ 对 μ 几乎处处的 ω 在 x_0 处均是方向可微的, 则上述命题的假设 (i) 与 (ii) 可推出公式 (5.277). 最后的假设 (iii) 是为保证 $\gamma'(x_0,\cdot)$ 的线性所需要的.

推论 5.109 设对每一 $\omega\in\Omega$, 函数 $g(\cdot,\omega)$ 是可微的, $g(x,\omega)$ 与 $\nabla_x g(x,\omega)$ 是连续的 (关于 x 与 ω 的联合变量), 则积分函数 $\gamma(x)$ 是连续可微的, 且对任何 $x_0\in\mathbb{R}^n$ 与任何 $\mu\in C(\Omega)^*$, 可交换性公式 (5.276) 成立.

证明 对测度 $\mu\in C(\Omega)^*$, 命题 5.108 中假设 (i) 由 $g(x,\cdot)$ 的连续性可得到, 假设 (ii) 与 (iii) 由 $\nabla_x g(x,\omega)$ 的连续性得到. 注意到, (5.275) 由中值定理得到, 其中相应的 Lipschitz 常数 $\kappa(\omega)$ 由 $\|\nabla_x g(x,\omega)\|$ 在 x_0 的邻域上取最大值得到. 所以, $\gamma(x)$ 是可微的, 公式 (5.276) 成立. 进一步, 由 Lebesgue 控制收敛定理可得 $\nabla\gamma(x)$ 是连续的. □

在 $\nabla_x g(x,\omega)$ 的连续性的假设条件下, 一阶最优性条件 (5.273) 可以表示为下述形式:

$$\nabla f(x_0) + \int_\Omega \nabla_x g(x_0,\omega)\mathrm{d}\mu(\omega) = 0, \quad \mathrm{supp}(\mu)\subset\Delta(x_0), \quad \mu\succeq 0. \tag{5.278}$$

引理 5.110 设存在测度 $\bar\mu\in C(\Omega)^*$ 满足条件 (5.273), 则存在离散测度 $\mu = \sum_{i=1}^m \lambda_i\delta(\omega_i)$ 满足 (5.274) 且 $m\leqslant n$.

证明 考虑满足 (5.273) 与 $\mu(\Omega)=\bar\mu(\Omega)$ 的所有测度 μ 的集合 $\mathcal{M}$. 集合 $\mathcal{M}$ 是非空的凸的弱 * 紧致的, 因而由 Krein-Milman 定理至少有一极点. 如我们以前证明

的, $\mathcal{M}$ 的任何极点均有至多含有 $n+1$ 个点的支撑. 还注意到 (5.274) 的第一个条件涉及 m 个未知量 λ_i 的 n 个方程. 所以, 若对某个 ω_i, 存在 λ_i, $i=1,\cdots,m$, 满足条件 (5.274), 则总可以找到这样的 λ_i, 其中仅有 n 个元素是非零的. □

现在设 $f(\cdot), g(\cdot,\omega), \omega\in\Omega$ 是连续可微的, $\nabla_x g(x,\omega)$ 是连续的. 可得相应的映射 $G(x)(\cdot):=g(x,\cdot)$ 是连续可微的且其微分由公式 (5.242) 给出. 由于集合 $K:=C_-(\Omega)$ 在空间 $C(\Omega)$ 中具有非空的内部, 这里的 Robinson 约束规范等价于下述条件: 存在 $h\in\mathbb{R}^n$ 满足

$$G(x_0)+DG(x_0)h\in \operatorname{int}(K). \tag{5.279}$$

结果, 上述条件等价于下述条件: 存在 $h\in\mathbb{R}^n$ 满足

$$h\cdot\nabla_x g(x_0,\omega)<0, \quad 对所有的\ \ \omega\in\Delta(x_0). \tag{5.280}$$

称上述条件 (5.280) 为广义的 Mangasarian-Fromovitz(MF) 约束规范.

定理 5.111　设 x_0 是 (不必要是凸的) 半无限规划问题 (P) 的局部最优解, $f(\cdot)$ 与 $g(\cdot,\omega)$, $\omega\in\Omega$ 是连续可微的, $\nabla_x g(x,\omega)$ 是连续的 (关于 x 与 ω 的联合变量), 则下述结论成立:

(i) 满足条件 (5.273) 的所有测度 $\mu\in C(\Omega)^*$ 的集合 $\Lambda(x_0)$ 非空有界的充分必要条件是广义的 Mangasarian-Fromovitz 约束规范(5.280) 成立.

(ii) 对任何 $m\geqslant n$, 若广义的 Mangasarian-Fromovitz 约束规范成立, 则满足 (5.274) 的所有离散测度 $\mu=\sum_{i=1}^{m}\lambda_i\delta(\omega_i)$ 的集合 $\Lambda_m(x_0)$ 是非空的有界的. 相反地, 若 $m\geqslant n+1$ 且集合 $\Lambda_m(x_0)$ 是非空的有界的, 则广义的 Mangasarian-Fromovitz 约束规范成立.

证明　在上述的可微性假设下, 映射 $G:\mathbb{R}^n\to C(\Omega)$ 是连续可微的, 且在当前情形之下, Robinson 约束规范与广义的 MF 约束规范是等价的, 由定理 3.9 得, 若广义的 MF 约束规范成立, 则集合 $\Lambda(x_0)$ 是非空的凸的有界的. 因为集合 $C_-(\Omega)$ 具有非空的内部, 由命题 3.17 得, 相反的结论也成立. 这证得结论 (i).

显然, $\Lambda_m(x_0)\subset\Lambda(x_0)$, 若 $\Lambda(x_0)$ 是有界的, 则 $\Lambda_m(x_0)$ 也是有界的. 由结论 (i), 广义 MF 约束规范可推出 $\Lambda(x_0)$ 是非空的, 因此, 由引理 5.110, 对任何 $m\geqslant n$, 集合 $\Lambda_m(x_0)$ 是非空的.

现在设 $\Lambda_m(x_0)$ 是非空有界的, 则 $\Lambda(x_0)$ 是非空的. 进一步, 对任何 $\bar\mu\in\Lambda(x_0)$, 假设 $m\geqslant n+1$, 可以找到离散测度 $\mu\in\Lambda_m(x_0)$ 满足 $\bar\mu(\Omega)=\mu(\Omega)$(见引理 5.110 的证明). 因为对 $\mu\succeq 0$, $\|\mu\|=\mu(\Omega)$, 可得 $\Lambda(x_0)$ 也是有界的. 结果, 由结论 (i) 可得广义 MF 约束规范成立. □

很有意思的是, 如下述例子说明, $m=n$ 时 $\Lambda_m(x_0)$ 的非空性与有界性不能推出广义 MF 约束规范.

例 5.112 考虑下述线性规划问题:

$$\min_{(x_1,x_2)\in\mathbb{R}^2} x_1 \quad \text{s. t.} \quad -x_2\leqslant 0,\quad -x_1+x_2\leqslant 0,\quad x_1+x_2\leqslant 0. \tag{5.281}$$

这一问题仅有一可行点 $x_0=(0,0)$, 它也是最优解. 不难看到, 对 (5.281) 中 (三个之中) 的任何两个约束的组合, 相应的 Lagrange 乘子的集合是有界的 (对某一组合它是空集), 因此 $\Lambda_2(x_0)$ 是有界的. 另一方面, 广义的 MF 约束规范是不成立的, 实际上, 这里的所有 Lagrange 乘子的集合是无界的.

因为集合 $C_-(\Omega)$ 具有非空内部, 由 3.1.2 节的一般性结果 (见命题 3.18) 可得, 对于半无限规划问题, 用广义 Lagrange 乘子刻画的最优性条件总是成立的, 直接推导此结果是具有指导性意义的. 首先观察到, 可行性约束 $g(x,\omega)\leqslant 0,\ \omega\in\Omega$ 可用一个约束 $\sup_{\omega\in\Omega} g(x,\omega)\leqslant 0$ 来代替. 令 x_0 是 (P) 的可行点, 考虑最大值函数

$$\phi(x):=\max\{f(x)-f(x_0),\ \sup_{\omega\in\Omega} g(x,\omega)\}. \tag{5.282}$$

显然, $\phi(x_0)=0$, 且若 x_0 是 (P) 的 (局部) 最优解, 则 x_0 是 $\phi(x)$ 在 $x\in\mathbb{R}^n$ 上的 (局部) 极小点. 置 $g(\cdot,\omega_0):=f(x)-f(x_0)$, $\overline{\Omega}:=\Omega\cup\{\omega_0\}$, 可以表示 $\phi(x)=\sup_{\omega\in\overline{\Omega}} g(x,\omega)$.

因为 $f(x)$ 与 $g(x,\omega)$ 是连续的, Ω 是紧致的, 有最大值函数 $\phi(x)$ 是连续的 (见命题 4.4). 进一步, 设 $f(\cdot)$ 与 $g(\cdot,\omega),\omega\in\Omega$ 是连续可微的, $\nabla_x g(x,\omega)$ 是连续的 (关于联合变量 x 与 ω). 由定理 4.13, $\phi(\cdot)$ 是 (Fréchet 意义下) 方向可微的, 且

$$\phi'(x_0,h):=\max_{\omega\in\overline{\Delta}(x_0)} h\cdot\nabla_x g(x_0,\omega), \tag{5.283}$$

其中 $\overline{\Delta}(x_0):=\{\omega\in\overline{\Omega}: g(x_0,\omega)=0\}$. 注意到, 因为 x_0 是可行的, 所以 $\phi(x_0)=0$, 集合 $\overline{\Delta}(x_0)$ 表示 $g(x_0,\cdot)$ 在 $\overline{\Omega}$ 上的极大值点的集合, 且 $\overline{\Delta}(x_0)=\{\omega_0\}\cup\Delta(x_0)$.

已经假设在 x_0 处起作用的约束集合 $\Delta(x_0)$ 是非空的, 否则, x_0 是 (P) 的可行集的内部点.

定理 5.113 *设 x_0 是 (不必为凸的) 半无限规划问题 (P) 的可行点, $f(\cdot)$ 与 $g(\cdot,\omega)$, $\omega\in\Omega$ 是连续可微的, $\nabla_x g(x,\omega)$ 是连续的 (关于联合变量 x 与 ω), 则下述结论成立:*

(i) *设 x_0 是 (P) 的局部最优解. 则存在不全为零的乘子 $\lambda_0,\cdots,\lambda_m$ 及点 $\omega_1,\cdots,\omega_m\in\Omega$ 满足 $m\leqslant n$, 且*

$$\lambda_0\nabla f(x_0)+\sum_{i=1}^{m}\lambda_i\nabla g(x_0,\omega_i)=0,\quad \omega_i\in\Delta(x_0),\quad \lambda_i\geqslant 0,\quad i=0,\cdots,m. \tag{5.284}$$

(ii) *相反地, 若条件*

$$0 \in \operatorname{int}\Big\{\operatorname{conv}(\nabla f(x_0), \bigcup_{\omega\in\Delta(x_0)} \nabla_x g(x_0,\omega))\Big\} \tag{5.285}$$

成立, 则 x_0 处的一阶增长条件成立, 因而 x_0 是 (P) 的严格局部最优解.

证明 若 x_0 是 (P) 的局部极小点, 则 x_0 是最大值函数 $\phi(x)$ 的 (无约束) 局部极小点. 对所有的 $h \in \mathbb{R}^n, \phi'(x_0,h) \geqslant 0$, 即方向导数函数 $\eta(\cdot) := \phi'(x_0,\cdot)$ 在 $h=0$ 处取极小值, 因此 $0 \in \partial\eta(0)$. 由公式 (5.283), 有函数 $\eta(\cdot)$ 是凸的, 且

$$\partial\eta(0) = \operatorname{con}\{\nabla f(x_0), \bigcup_{\omega\in\Delta(x_0)} \nabla_x g(x_0,\omega)\}. \tag{5.286}$$

注意到, 由于 $\Delta(x_0)$ 是紧致集, $\nabla_x g(x_0,\cdot)$ 是连续的, 上述等式右端的集合是紧致的, 因而是闭的. 由 (5.286), 条件 $0 \in \partial\eta(0)$ 等价于 (5.284). 这完成了结论 (i) 的证明.

若 $0 \in \operatorname{int}[\partial\eta(0)]$, 即如果条件 (5.285) 成立, 则对所有的 $h \neq 0$, $\phi'(x_0,h) > 0$. 由于 $\phi(\cdot)$ 在 x_0 处是 Fréchet 意义下方向可微的, 可推出存在 $c > 0$, 满足 $\phi(x) \geqslant c\|x - x_0\|$ 对 x_0 的某一邻域中所有 x 均成立, 可推出问题 (P) 在点 x_0 处的一阶增长条件是成立的. 当然, 由一阶增长条件可推出 x_0 的严格局部最优性. □

若广义的 Mangasarian-Fromovitz 约束规范成立, 则一阶必要条件 (5.284) 中乘子 λ_0 非零. 因此, 条件 (5.284) 变成与 (5.274) 相同的条件.

讨论满足一阶最优性条件的 Lagrange 乘子测度唯一性的问题. 由引理 5.110 可观察到, 若满足最优性条件 (5.273) 的测度是唯一的, 则它应该是离散测度.

定理 5.114 令 $\mu := \sum_{i=1}^{m} \lambda_i \delta(\omega_i)$ 是离散测度, $\lambda_i > 0, i = 1,\cdots,m$ 满足最优性条件 (5.274), 则 μ 是唯一的当且仅当下述两个条件成立:

(i) 梯度向量 $\nabla_x g(x_0,\omega_i)$, $i = 1,\cdots,m$ 是线性无关的.

(ii) 对集合 $\{\omega_1,\cdots,\omega_m\}$ 的任何邻域 $N \subset \Omega$, 存在 $h \in \mathbb{R}^n$ 满足

$$h \cdot \nabla_x g(x_0,\omega_i) = 0, \quad i = 1,\cdots,m, \tag{5.287}$$

$$h \cdot \nabla_x g(x_0,\omega) < 0, \quad \omega \in \Delta(x_0) \setminus N. \tag{5.288}$$

证明 假设上述假设 (i) 与 (ii) 成立, 令 μ' 是满足最优性条件 (5.273) 的另一测度. 由 Krein-Milman 定理有, 若 μ' 不同于 μ, 则可以找到满足 (5.273) 且不同于 μ 的离散测度. 所以可假设 $\mu' := \sum_{i=1}^{k} \lambda'_i \delta(\omega'_i)$ 是离散的, 其支撑是 $\{\omega'_1,\cdots,\omega'_k\}$. 如果 μ' 的支撑包含在集合 $\{\omega_1,\cdots,\omega_m\}$ 中, 则由线性无关性条件 (i) 得 $\mu' = \mu$. 如果 μ' 的支撑不包含在集合 $\{\omega_1,\cdots,\omega_m\}$ 中, 则存在 $\{\omega_1,\cdots,\omega_m\}$ 的邻域 N 满足 $\omega'_i \notin N$, $i \in \mathcal{I}$, 其中 $\mathcal{I} := \{i : \omega'_i \notin \{\omega_1,\cdots,\omega_m\},\ i = 1,\cdots,k\}$. 令 h 是满足 (5.287)

与 (5.288) 的相应的向量. 有

$$h\cdot[\nabla_x L(x_0,\mu')-\nabla_x L(x_0,\mu)]=\sum_{i\in\mathcal{I}}\lambda_i'[h\cdot\nabla_x g(x_0,\omega_i')].$$

因为 $\lambda_i'>0$ 且 $h\cdot\nabla_x g(x_0,\omega_i')<0, i\in\mathcal{I}$, 得到上述方程的右端是负的. 然而, 这同最优性条件 $\nabla_x L(x_0,\mu)=0$ 与 $\nabla_x L(x_0,\mu')=0$ 是相矛盾的.

相反地, 设测度是满足条件 (5.273) 的唯一测度. 显然, 可推出假设 (i) 中的线性无关性条件. 现在设假设 (ii) 是不成立的, 则存在集合 $\{\omega_1,\cdots,\omega_m\}$ 的开邻域 N 满足

$$\{h:h\cdot\nabla_x g(x_0,\omega_i)=0,\ i=1,\cdots,m,\ \ h\cdot\nabla_x g(x_0,\omega)<0,\ \omega\in\Xi\}=\varnothing, \tag{5.289}$$

其中 $\Xi:=\Delta(x_0)\setminus N$. 因为 Ξ 是紧致集 Ω 中的闭子集, 集合 Ξ 亦是紧致集合. 考虑下述映射:

$$\mathcal{G}(x):=((g(x,\omega_1),\cdots,g(x,\omega_m)),g(x,\cdot)):\mathbb{R}^n\to\mathbb{R}^m\times C(\Xi).$$

可得 $\mathcal{G}(x_0)=0$, 且条件 (5.289) 意味着映射 $\mathcal{G}$ 在 x_0 处关于集合 $\mathcal{K}:=\{0\}\times C_+(\Xi)$ 的 Robinson 约束规范是不成立的. 由于集合 $C_+(\Xi)$ 在空间 $C(\Xi)$ 中具有非空的内部, 由命题 2.97 得上述 Robinson 约束规范等价于条件 (2.183). 在当前情况, 条件 (2.183) 意味着若 $(\alpha_1,\cdots,\alpha_m)\in\mathbb{R}^m, \hat{u}\in C(\Omega)^*$ 满足

$$\sum_{i=1}^m\alpha_i\nabla_x g(x_0,\omega_i)=0,\quad \int_\Omega\nabla_x g(x_0,\omega)\mathrm{d}\hat{\mu}(\omega)=0,\quad \mathrm{supp}(\hat{\mu})\subset\Xi,\quad \hat{\mu}\succeq 0, \tag{5.290}$$

则 $\alpha_i=0, i=1,\cdots,m$ 且 $\hat{\mu}=0$. 因此, 由 (5.289), 上述约束规范不成立, 则存在非零的 $((\alpha_1,\cdots,\alpha_m),\hat{\mu})$ 满足 (5.290). 考虑相应的测度 $\mu':=\sum_{i=1}^m\alpha_i\delta(\omega_i)+\hat{\mu}$. 有 $\mu'\neq 0$, 且对 $t>0$ 充分小, 测度 $\mu+t\mu'$ 满足最优性条件 (5.273), 这与 μ 的唯一性是矛盾的. 证毕. □

定理 5.114 条件 (ii) 中的向量 h 一般来说是依赖于邻域 N 的. 一个很自然的问题是这个条件 (ii) 是否可以被下述更强的条件代替:

(ii′) 存在 $h\in\mathbb{R}^m$ 满足

$$h\cdot\nabla_x g(x_0,\omega_i)=0,\quad i=1,\cdots,m, \tag{5.291}$$

$$h\cdot\nabla_x g(x_0,\omega)<0,\quad \omega\in\Delta(x_0)\backslash\{\omega_1,\cdots,\omega_m\}. \tag{5.292}$$

不难看到, 若起作用约束的集合 $\Delta(x_0)$ 是有限的, 则条件 (ii) 与 (ii′) 是等价的. 然而, 如下述例子所示, 通常而言, 条件 (ii′) 对于 Lagrange 乘子测度 μ 的唯一性并不是必要的.

例 5.115 令 $\Omega := [0,4] \subset \mathbb{R}$, 考虑 $g : \mathbb{R}^3 \times [0,4] \to \mathbb{R}$ 具有形式 $g(x,\omega) = x_1a_1(\omega) + x_2a_2(\omega) + x_3a_3(\omega)$, 其中

$$a_1(\omega) := \begin{cases} \omega^2, & \omega \in [0,1], \\ 1.5 - 0.5\omega, & \omega \in [1,3], \\ 0, & \omega \in [3,4], \end{cases}$$

$$a_2(\omega) := \begin{cases} -\omega, & \omega \in [0,1], \\ \omega - 2, & \omega \in [1,4], \end{cases}$$

且 $a_3(\omega) := 1, \omega \in [0,4]$. 令 $f(x)$ 是线性函数, 满足 $\nabla f(x) = (0,0,-1)$. 有 $x_0 := 0$, $\nabla f(x_0) + \nabla_x g(x_0, 0) = 0$, 且 $\Delta(x_0) = [0,4]$. 所以, 在 $x_0 = 0$ 处的一阶最优性条件成立, Lagrange 乘子测度 $\mu = \delta(\omega_1)$, $\omega_1 = 0$, 由于规划问题是凸的, $x_0 = 0$ 是问题的最优解. 还有, 对于 $h := (0,0,-1)$ 及所有的 $\omega \in [0,4]$, $h \cdot \nabla_x g(x_0,\omega) = -1$, 因此广义 Mangasarian-Fromovitz 约束规范成立.

这里条件 (ii′) 是不成立的. 事实上, 假设存在向量 $h = (h_1, h_2, h_3)$ 满足 (5.291) 与 (5.292). 由 (5.291) 可得 $h_3 = 0$, 由 (5.292) 可得 $h_2 < 0$. 得到 $h \cdot \nabla_x g(x_0, 0) = 0$ 且 $\partial[h \cdot \nabla_x g(x_0, 0)]/\partial\omega > 0$. 因此, 对充分小的 $\omega > 0$, $h \cdot \nabla_x g(x_0, \omega)$ 是正的, 这当然与 (5.292) 是矛盾的.

另一方面, 不难验证, 在这里, 定理 5.114 中条件 (i) 与 (ii) 是成立的, 因此 μ 是唯一的. 这说明条件 (ii) 与 (ii′) 不是等价的, 且条件 (ii′) 对 μ 的唯一性不是必要的.

5.4.3 二阶最优性条件

这一节讨论 (不必是凸的) 半无限规划问题 (5.241) 的二阶必要和充分性条件. 设集合 Ω 是空间 $\mathbb{R}^q$ 的紧致子集, 函数 $f : \mathbb{R}^n \to \mathbb{R}$ 与 $g : \mathbb{R}^n \times \mathbb{R}^q \to \mathbb{R}$ 是二次连续可微的, 可以用第 3 章中的一般性结果写出半无限规划问题的二阶最优性条件. 然而以一种更直接的方式来推导这样的二阶条件是具有指导意义的, 这一更直接的方式通过最优值函数:

$$v(x) := \sup_{\omega \in \Omega} g(x, \omega) \tag{5.293}$$

与 (5.282) 中定义的最大值函数 $\phi(x)$ 来实现, $\phi(x)$ 也可以表示为

$$\phi(x) = \max\{f(x) - f(x_0), v(x)\}.$$

将 (5.293) 右端的优化问题称为下层优化问题.

令 x_0 是问题 (P) 的可行点. 已经假定集合 $\Delta(x_0)$ 是非空的. 此种情况, $\Delta(x_0)$ 表示 $g(x_0, \cdot)$ 在 Ω 上的最大值点的集合, 即 $\Delta(x_0)$ 是下层问题的最优解集合.

(P) 在点 x_0 处的临界锥可以表示为如下形式:

$$C(x_0) = \{h \in \mathbb{R}^n : h \cdot \nabla_x g(x_0, \omega) \leqslant 0,\ \omega \in \Delta(x_0),\ h \cdot \nabla f(x_0) \leqslant 0\}. \tag{5.294}$$

由 (5.283) 可得

$$\phi'(x_0, h) = \max\left\{h \cdot \nabla f(x_0),\ \sup_{\omega \in \Delta(x_0)} h \cdot \nabla_x g(x_0, \omega)\right\}. \tag{5.295}$$

所以, $C(x_0) = \{h:\ \phi'(x_0, h) \leqslant 0\}$. 若进一步, 点 x_0 是稳定的, 即存在满足一阶最优性条件 (5.273) 的测度 $\bar{\mu}$, 则 (见命题 3.10)

$$C(x_0) = \{h:\ h \cdot \nabla_x g(x_0, \omega) \leqslant 0,\ \omega \in \Delta(x_0),\ h \cdot \nabla f(x_0) = 0\}, \tag{5.296}$$

或等价地

$$\begin{aligned} C(x_0) = & \Big\{h : h \cdot \nabla_x g(x_0, \omega) \leqslant 0,\ \omega \in \Delta(x_0), \\ & \int_\Omega h \cdot \nabla_x g(x_0, \omega) \mathrm{d}\bar{\mu}(\omega) = 0\Big\} \\ = & \{h : h \cdot \nabla_x g(x_0, \omega) = 0, \omega \in \operatorname{supp}(\bar{\mu}) \\ & h \cdot \nabla_x g(x_0, \omega) \leqslant 0, \omega \in \Delta(x_0) \backslash \operatorname{supp}(\bar{\mu})\}. \end{aligned}$$

由 (5.294) 与 (5.295) 可得, $C(x_0) = \{0\}$ 当且仅当对任何 $h \neq 0$, 不等式 $\phi'(x_0, h) > 0$ 成立. 结果这与条件 (5.285) 是等价的. 因此, 若 $C(x_0) = \{0\}$, 则由定理 5.113, 问题 (P) 在 x_0 处的一阶增长条件是成立的. 若 x_0 是 (P) 的稳定点且广义的 MF 约束规范成立, 则一阶增长条件在 x_0 处成立的充分必要条件是 $C(x_0) = \{0\}$.

因为在后续的分析中处理内二阶切集, 为简便起见, 在这一节假设集合 Ω 在每一点 $\omega \in \Delta(x_0)$ 处的内切锥与外切锥是重合的, 即

$$T_\Omega(\omega) = T_\Omega^i(\omega), \quad \forall\, \omega \in \Delta(x_0). \tag{5.297}$$

若集合 Ω 可表示为有限多个凸集的并, 则这一假设是自动成立的.

考虑一点 $\overline{\omega} \in \Delta(x_0)$. 由于 $\overline{\omega}$ 是函数 $g(x_0, \cdot)$ 在 Ω 上的最大值点, 由下层问题的一阶必要条件可得

$$\eta \cdot \nabla_\omega g(x_0, \overline{\omega}) \leqslant 0, \quad \forall \eta \in T_\Omega(\overline{\omega}). \tag{5.298}$$

下层问题相应的临界锥可表示为

$$\mathcal{C}(\overline{\omega}) := \{\eta \in T_\Omega(\overline{\omega}):\ \eta \cdot \nabla_\omega g(x_0, \overline{\omega}) = 0\}. \tag{5.299}$$

对给定的 $h \in \mathbb{R}^n$, 考虑最优化问题

$$\max_{\eta \in \mathcal{C}(\overline{\omega})} \left\{Q_{\overline{\omega}}(h, \eta) + \sigma(\nabla_\omega g(x_0, \overline{\omega}), T_\Omega^{i,2}(\overline{\omega}, \eta))\right\}, \tag{5.300}$$

其中

$$Q_{\overline{\omega}}(h,\eta) := 2h \cdot \nabla^2_{x\omega} g(x_0,\overline{\omega})\eta + \eta \cdot \nabla^2_{\omega\omega} g(x_0,\overline{\omega})\eta.$$

用 $\vartheta(\overline{\omega},h)$ 记上述问题 (5.300) 的最优值, 置

$$\theta(\overline{\omega},h) := h \cdot \nabla^2_{xx} g(x_0,\overline{\omega})h + \vartheta(\overline{\omega},h). \tag{5.301}$$

注意到, 对 $\eta = 0, Q_{\overline{\omega}}(h,\eta) = 0$, (5.300) 中相应的 sigma 项是零, 因此

$$\vartheta(\overline{\omega},h) \geqslant 0. \tag{5.302}$$

由命题 4.129, 对任何 $h \in \mathbb{R}^n$, 有

$$\liminf_{\substack{t \downarrow 0 \\ h' \to h}} \frac{v(x_0+th') - v(x_0) - th' \cdot \nabla_x g(x_0,\overline{\omega})}{\frac{1}{2}t^2} \geqslant \theta(\overline{\omega},h). \tag{5.303}$$

这导致下述的二阶充分性条件. 令

$$L^g(x,\lambda_0,\mu) := \lambda_0 f(x) + \int_\Omega g(x,\omega)\mathrm{d}\mu(\omega)$$

是半无限问题 (P) 的广义 Lagrange 函数. 离散测度 $\mu = \sum_{i=1}^m \lambda_i \delta(\omega_i)$ 具有形式:

$$L^g(x,\lambda_0,\mu) = \lambda_0 f(x) + \sum_{i=1}^m \lambda_i g(x,\omega_i).$$

定理 5.116　令 x_0 是半无限规划问题 (P) 的可行点, 满足一阶必要性条件 (5.284). 设下述二阶条件成立: 对每一 $h \in C(x_0)\backslash\{0\}$, 存在满足条件 (5.284) 的乘子 $\lambda_0, \mu = \sum_{i=1}^m \lambda_i \delta(\omega_i)$, 使得

$$h \cdot \nabla^2_{xx} L^g(x_0,\lambda_0,\mu)h + \sum_{i=1}^m \lambda_i \vartheta(\omega_i,h) > 0. \tag{5.304}$$

则二阶增长条件在点 x_0 处成立.

证明　只需证明, 存在某一 $c > 0$ 满足对 x_0 的邻域中所有的 x, 有 $\phi(x) \geqslant c\|x-x_0\|^2$. 注意到 $\phi(x_0) = 0$, 因为假设 $\Delta(x_0) \neq \varnothing$, 有 $v(x_0) = 0$. 用反证法. 假设这是不真的, 则存在序列 $h_k \in \mathbb{R}^n, t_k \downarrow 0$, 满足 $\|h_k\| = 1$ 且

$$\phi(x_0 + t_k h_k) \leqslant o(t_k^2). \tag{5.305}$$

如有必要, 可取一子列, 不妨设 $h_k \to \bar{h}$, 其中 $\bar{h}$ 满足 $\|\bar{h}\| = 1$. 由于 $\phi(x)$ 在 x_0 处是 Hadamard 方向可微的, 由 (5.4.3)[①]得 $\phi'(x_0, \bar{h}) \leqslant 0$. 因此, 条件 (5.284) 意味着 $\phi'(x_0, h) \geqslant 0$ 对所有 $h \in \mathbb{R}^n$ 均成立. 结果, $\phi'(x_0, \bar{h}) = 0$, 即 $\bar{h} \in C(x_0)$.

令 λ_0 与 $\mu = \sum_{i=1}^{m} \lambda_i \delta(\omega_i)$ 是满足相应的不等式 (5.304) 且 $\lambda_0 + \sum_{i=1}^{m} \lambda_i = 1$ 的乘子, 得到

$$\phi(\cdot) \geqslant \lambda_0(f(\cdot) - f(x_0)) + \sum_{i=1}^{m} \lambda_i v(\cdot). \tag{5.306}$$

由于 $f(x)$ 是二次连续可微的, 有

$$f(x_0 + t_k h_k) - f(x_0) = t_k h_k \cdot \nabla f(x_0) + \frac{1}{2} t_k^2 \bar{h} \cdot \nabla^2 f(x_0) \bar{h} + o(t_k^2),$$

由 (5.303) 及对任何 $i \in \{1, \cdots, m\}$,

$$v(x_0 + t_k h_k) \geqslant t_k h_k \cdot \nabla_x g(x_0, \omega_i) + \frac{1}{2} t_k^2 \theta(\omega_i, \bar{h}) + o(t_k^2).$$

连同 (5.306), 上述的表达式可推出

$$\phi(x_0 + t_k h_k) \geqslant t_k h_k \cdot \nabla_x L^g(x_0, \lambda_0, \mu) + \frac{1}{2} t_k^2 [\bar{h} \cdot \nabla_{xx}^2 L^g(x_0, \lambda_0, \mu) \bar{h} + \sum_{i=1}^{m} \lambda_i \vartheta(\omega_i, \bar{h})] + o(t_k^2).$$

由一阶最优性条件 (5.284) 得到 $\nabla_x L^g(x_0, \lambda_0, \mu) = 0$. 将上述不等式与 (5.304) 相结合. 得到存在某一 κ, 有

$$\phi(x_0 + t_k h_k) \geqslant \kappa t_k^2 + o(t_k^2).$$

然而, 这与 (5.305) 是矛盾的. □

若 Ω 是多面体的, 则 (5.300) 中的 sigma 项会消失. 在任何情况下, 由 (5.302) 可得, 条件 (5.304) 比下述条件:

$$h \cdot \nabla_{xx}^2 L^g(x_0, \lambda_0, \mu) h > 0. \tag{5.307}$$

要更弱. 还注意到, 如果广义 Mangasarian-Fromovitz 约束规范(5.280) 成立, 则可取 $\lambda_0 = 1$, 在上述定理中可以在二阶充分性条件中用 Lagrange 乘子 (代替广义乘子).

若 $\overline{\omega} \in \Delta(x_0)$ 是 Ω 的孤立点, 则 $\mathcal{C}(\overline{\omega}) = \{0\}$, 因此 $\vartheta(\overline{\omega}, h) = 0$. 所以, 若集合 Ω 是有限的, 条件 (5.304) 简化为条件 (5.307). 若 $\overline{\omega} \in \Delta(x_0)$ 是 Ω 的内点, 则 $\mathcal{C}(\overline{\omega}) = \mathbb{R}^q$, 从而

$$\vartheta(\overline{\omega}, h) = \sup_{\eta \in \mathbb{R}^q} \left\{ 2h \cdot \nabla_{x\omega}^2 g(x_0, \overline{\omega}) \eta + \eta \cdot \nabla_{\omega\omega}^2 g(x_0, \overline{\omega}) \eta \right\}.$$

① 应该是 (5.305)

若进一步, Hesse 阵 $\nabla^2_{\omega\omega}g(x_0,\overline{\omega})$ 是非奇异的, 因此它是负定的, 则

$$\vartheta(\overline{\omega},h)=-h[\nabla^2_{x\omega}g(x_0,\overline{\omega})][\nabla^2_{\omega\omega}g(x_0,\overline{\omega})]^{-1}[\nabla^2_{x\omega}g(x_0,\overline{\omega})]^{\mathrm{T}}h.$$

另一方面, 若 $\nabla^2_{\omega\omega}g(x_0,\overline{\omega})$ 是奇异的, 则 $\vartheta(\overline{\omega},h)$ 可取 $+\infty$ 值. 若存在 η 满足 $\eta\cdot\nabla^2_{\omega\omega}g(x_0,\overline{\omega})\eta=0$, 而 $\eta\cdot\nabla^2_{x\omega}g(x_0,\overline{\omega})\eta\neq 0$, 这种情况是可能发生的.

注 5.117　设集合 Ω 由

$$\Omega:=\{\omega\in\mathbb{R}^q:\ \mathcal{G}(\omega)\in\mathcal{K}\}\tag{5.308}$$

给出, 其中 $\mathcal{K}$ 是 Banach 空间Z 的闭凸子集, $\mathcal{G}:\mathbb{R}^q\to Z$ 是二次连续可微的映射. 记

$$\mathcal{L}(x,\omega,\alpha):=g(x,\omega)-\langle\alpha,\mathcal{G}(\omega)\rangle,\quad \alpha\in Z^*$$

为下层问题的 Lagrange 函数. 令 $\overline{\omega}\in\Delta(x_0),\mathcal{A}(x_0,\overline{\omega})$ 是满足下述最优性条件的相应的 Lagrange 乘子集合:

$$\nabla_\omega\mathcal{L}(x_0,\overline{\omega},\alpha)=0,\quad \alpha\in N_{\mathcal{K}}(\mathcal{G}(\overline{\omega})).\tag{5.309}$$

若下层问题在 $\overline{\omega}$ 处 Robinson 约束规范成立, 则集合 $\mathcal{A}(x_0,\overline{\omega})$ 是非空的有界的. 进一步, 由二阶切集的链式法则 (见命题 3.33) 可得, 在问题 (5.300) 中括弧里的目标函数值可以表示为下述的最优化问题:

$$\begin{aligned}\max_{w\in\mathbb{R}^q}\quad & w\cdot\nabla_\omega g(x_0,\overline{\omega})+Q_{\overline{\omega}}(h,\eta),\\ \text{s. t.}\quad & D\mathcal{G}(\overline{\omega})w+D^2\mathcal{G}(\overline{\omega})(\eta,\eta)\in\mathcal{T}^2_{\mathcal{K}}(\eta),\end{aligned}$$

其中

$$\mathcal{T}^2_{\mathcal{K}}(\eta):=T^{i,2}_{\mathcal{K}}(\mathcal{G}(\overline{\omega}),D\mathcal{G}(\overline{\omega})\eta).$$

通过计算上述问题的对偶可得 (见注 4.132), (5.301) 定义的值 $\theta(\overline{\omega},h)$ 由下述极大-极小问题的最优值给出:

$$\max_{\eta\in\mathcal{C}(\overline{\omega})}\ \min_{\alpha\in\mathcal{A}(x_0,\overline{\omega})}\left\{D^2\mathcal{L}(x_0,\overline{\omega},\alpha)((h,\eta),(h,\eta))+\sigma(\alpha,\mathcal{T}^2_{\mathcal{K}}(\eta))\right\}.\tag{5.310}$$

注意到, 二阶充分性条件 (5.304) 可以表示为下述等价的形式:

$$\lambda_0[h\cdot\nabla^2 f(x_0)h]+\sum_{i=1}^m\lambda_i\theta(\omega_i,h)>0.\tag{5.311}$$

若集合 $\mathcal{K}$ 是多面体的, 即下层问题是非线性规划问题, 则 (5.310) 中的 sigma 项消失. 进一步, 如果 $\mathcal{A}(x_0,\overline{\omega})$ 是单点集, 即下层问题有唯一的 Lagrange 乘子向量, 则

问题 (5.310) 成为二次规划问题. 若进一步, 临界锥 $\mathcal{C}(\bar{\omega})$ 是线性空间, 则 (5.310) 的最优值可以写为封闭的形式.

一个很自然的问题是定理 5.116 的二阶充分性条件是否是 "无间隙" 条件, 即相应的二阶必要性条件可通过将 (5.304) 中的严格不等号代替为符号 "$\geqslant$" 而得到. 为确保这样的 "无间隙" 性质, 需要额外的假设.

若最优值函数是 $v(x)$, 因而最大值函数 $\phi(x)$ 在 x_0 处是二阶方向可微的, 则相应的二阶必要性条件可以表示为 (见命题 3.105)

$$\phi''(x_0;h,w)\geqslant 0,\quad \forall h\in C(x_0),\quad \forall w\in\mathbb{R}^n. \tag{5.312}$$

设集合 $\Delta(x_0)$ 是有限的, 如 $\Delta(x_0)=\{\omega_1,\cdots,\omega_p\}$, 且 Ω 在集合 $\Delta(x_0)$ 的每一点处均是二阶正则的. 进一步, 设下层问题的二阶增长条件成立, 即对每一 $\overline{\omega}\in\Delta(x_0)$, 存在 $c>0$ 与 $\overline{\omega}$ 的邻域 $\mathcal{N}$ 满足

$$-g(x_0,\omega)\geqslant c\|\omega-\overline{\omega}\|^2,\quad \forall\omega\in\Omega\cap\mathcal{N}. \tag{5.313}$$

因为集合 Ω 是紧致的, 对任何 $x\in\mathbb{R}^n$, 下层问题有最优解, 由于上述的二阶增长条件, 这样的最优解是 Lipschitz 稳定的. 由定理 4.133 可得 $v(x)$ 在 x_0 处是二阶方向可微的, 且

$$v''(x_0;h,w)=\max_{\omega\in\Delta_1(x_0,h)}\left\{w\cdot\nabla_x g(x_0,\omega)+\theta(\omega,h)\right\}, \tag{5.314}$$

其中 $\theta(\omega,h)$ 由 (5.301) 定义, 且

$$\Delta_1(x_0,h):=\operatorname*{argmax}_{\omega\in\Delta(x_0)} h\cdot\nabla_x g(x_0,\omega). \tag{5.315}$$

这导致下述的二阶必要性条件.

定理 5.118 令 x_0 是半无限规划问题 (P) 的局部最优解, 设 (在 x_0 处起作用的约束) 集合 $\Delta(x_0)$ 是有限的, 集合 Ω 在集 $\Delta(x_0)$ 的每一点均是二阶正则的, 且对于下层问题, 二阶增长条件成立, 则下述二阶必要性条件是成立: 对每一 $h\in C(x_0)$, 存在乘子 λ_0, $\mu=\sum_{i=1}^{m}\lambda_i\delta(\omega_i)$, 不全为零, 满足一阶必要性条件 (5.284), 且

$$h\cdot\nabla_{xx}^2 L^g(x_0,\lambda_0,\mu)h+\sum_{i=1}^{m}\lambda_i\vartheta(\omega_i,h)\geqslant 0. \tag{5.316}$$

证明 这里用与定理 5.116 的证明中一样的记号. 根据 (5.314), 二阶必要性条件 (5.312) 可以表示为下述的形式: 对每一 $h\in C(x_0)$, 下述问题:

$$\min_{z,w} z\quad \text{s. t. }\ w\cdot\nabla_x g(x_0,\omega)+\theta(\omega,h)\leqslant z,\quad \omega\in\bar{\Delta}(x_0)$$

的最优值是非负的. 这是相容的线性规划问题, 因此这一问题与其对偶问题间不存在对偶间隙. 通过计算相应的对偶得到, 这一问题的最优值由满足 $\lambda_0+\cdots+\lambda_m=1$ 的所有广义 Lagrange 乘子$\lambda_0,\cdots,\lambda_m$ 取 (5.316) 左端的极大化给出. 证毕. □

在上述定理的条件之下, 得到了分别由定理 5.118 与 5.116 给出的一对 "无间隙" 二阶必要性与二阶充分性条件.

注 5.119 (5.314) 成立, 从而 (5.304) 与 (5.316) 分别是一对 "无间隙" 二阶充分条件与二阶必要性条件的另一种情况是当集合 Ω 由有限多个光滑约束定义:

$$\Omega:=\{\omega\in\mathbb{R}^q:\ \gamma_i(\omega)=0,\ i=1,\cdots,k,\ \gamma_i(\omega)\leqslant 0,\ i=k+1,\cdots,l\},$$

且下层最优化问题满足定理 4.142 假设情况, 即对下层问题, 在每一点 $\overline{\omega}\in\Delta(x_0)$ 处线性无关条件成立. 进一步, 对每一 $h\in C(x_0)$ 与每一 $\overline{\omega}\in\Delta(x_0,h)$①, 在 $\overline{\omega}$ 处严格互补条件成立, 集合 $\Delta(x_0)$ 是 $\overline{\omega}$ 的某一邻域内的光滑流形, 相应的二阶充分性条件 (等价于二阶增长性条件) 在 $\overline{\omega}$ 处是成立的. 注意到, 在这种情形下, 对每一 $\overline{\omega}\in\Delta(x_0,h)$②,

$$\theta(\overline{\omega},h)=\max_{\eta\in\mathcal{C}(\overline{\omega})} D^2\mathcal{L}(x_0,\overline{\omega},\alpha(\overline{\omega}))((h,\eta),(h,\eta)) \tag{5.317}$$

且

$$\mathcal{C}(\overline{\omega})=\{\eta:\ \eta\cdot\nabla\gamma_i(\overline{\omega})=0,\ i\in\{i,\cdots,k\}\cup I(\overline{\omega})\},$$

其中 $I(\overline{\omega})$ 是 $\overline{\omega}$ 处起作用的不等式约束的指标集,

$$\mathcal{L}(x,\omega,\alpha):=g(x,\omega)-\sum_{i=1}^{l}\alpha_i\gamma_i(\omega)$$

是下层问题的 Lagrange 函数, $\alpha(\omega)$ 是相应的 (在线性无关约束规范下, 它是唯一的)Lagrange 乘子向量.

上述二阶最优性条件可以与第 3 章讨论的抽象的二阶条件相比较. 由定理 4.148, 对任何 $y\in K:=C_-(\Omega)$ 与 $z\in T_K(y)$, 存在函数 $\tau_{y,z}:\Omega\to\overline{\mathbb{R}}$ 满足

$$T_K^{i,2}(y,z)=\{\psi\in C(\Omega):\ \psi(\omega)\leqslant\tau_{y,z}(\omega),\omega\in\Omega\}. \tag{5.318}$$

注意到, 由定理 4.149, 对任何满足 $y(\omega)\neq 0$ 的 ω, 有 $\tau_{y,z}(\omega)=+\infty$. 因此, 对 $\mu=\sum_{i=1}^{m}\lambda_i\delta(\omega_i)$ 与 $h\in\mathbb{R}^n$, 抽象二阶条件出现的 "sigma 项" 变为

$$\sigma(\mu,T_K^{i,2}(G(x_0),DG(x_0)h)=\sum_{i=1}^{m}\lambda_i\tau_{y,z}(\omega_i), \tag{5.319}$$

① 应为 $\overline{\omega}\in\Delta_1(x_0,h)$.

② 应为 $\overline{\omega}\in\Delta_1(x_0,h)$.

其中 $y(\cdot) := g(x_0, \cdot), z(\cdot) := h \cdot \nabla_x g(x_0, \cdot)$.

一般而言, 上述函数 $\tau_{y,z}(\cdot)$ 的计算并不是一个平凡的工作. 由 4.10.3 节的结果 (见命题 4.158 与 4.159) 得到, 在定理 5.118 的假设之下, $\tau_{y,z}(\cdot)$ 是问题 (5.300) 的最优值, 即

$$\tau_{y,z}(\omega) = \vartheta(\omega, h), \quad \forall \omega \in \Delta(x_0). \tag{5.320}$$

此种情况下, 二阶最优性条件(5.304) 与 (5.316) 中的项 $\sum_{i=1}^{m} \lambda_i \vartheta(\omega_i, h)$ 与 (5.319) 给出的 sigma 项是重合的.

5.4.4 扰动性分析

这一节讨论半无限规划问题的扰动分析. 考虑参数化问题:

$$(\mathrm{P}_u) \qquad \min_{x \in \mathbb{R}^n} f(x, u) \ \text{ s. t. } \ g(x, u, \omega) \leqslant 0, \quad \forall \omega \in \Omega, \tag{5.321}$$

它依赖于参数向量 $u \in U$. 设参数集合 U 是 Banach 空间, Ω 是紧致的度量空间, 函数 $f(x,u)$ 与 $g_\omega(x,u) := g(x,u,\omega), \omega \in \Omega$ 在 $\mathbb{R}^n \times U$ 上是连续可微的, 且 $Dg_\omega(x,u)$ 在 $\mathbb{R}^n \times U \times \Omega$ 上是连续的. 注意到, 可以表示为

$$Dg_\omega(x_0, u_0)(h, d) = h \cdot \nabla_x g_\omega(x_0, u_0) + D_u g_\omega(x_0, u_0) d,$$

类似可得 $Df(x,u)(h,d)$ 的表达式. 还设对参数向量的特定值 u_0, 相应的问题 (P_{u_0}) 与非扰动问题 (P) 是重合的, 即 $f(\cdot, u_0) = f(\cdot)$ 且 $g(\cdot, u_0, \cdot) = g(\cdot, \cdot)$. 用 $\Phi(u)$ 记 (P_u) 的可行集合, 用 $v(u)$ 记 (P_u) 的最优值, 用 $\mathcal{S}(u)$ 记 (P_u) 的最优解集合.

称下确界–紧致性条件成立, 若存在 $\alpha \in \mathbb{R}$ 满足: 对 u_0 的某一邻域中所有的 u, 水平集 $\{x \in \Phi(u) : \ f(x,u) \leqslant \alpha\}$ 是非空的一致有界的. 此种情形, 对接近 u_0 的所有的 u, (P_u) 有最优解 $\bar{x}(u)$, 假设广义的 Mangasarian-Fromovitz 约束规范在 (P) 的某最优解处成立 (见命题 4.4), 当 $u \to u_0$ 时 $\mathrm{dist}(\bar{x}(u), \mathcal{S}(u_0)) \to 0$.

令 $x_0 \in \mathcal{S}(u_0)$ 是 (P) 的最优解, 令

$$\Delta(x_0, u_0) := \{\omega \in \Omega : \ g(x_0, u_0, \omega) = 0\}$$

是相应的紧约束集合. 对给定的方向 $d \in U$, 考虑 (P_u) 的下述线性化:

$$(\mathrm{PL}_d) \qquad \begin{array}{ll} \min\limits_{h \in \mathbb{R}^n} & Df(x_0, u_0)(h, d), \\ \text{s. t.} & Dg_\omega(x_0, u_0)(h, d) \leqslant 0, \quad \forall \omega \in \Delta(x_0, u_0). \end{array} \tag{5.322}$$

这是线性的半无限规划问题. 它的对偶问题是

$$(\mathrm{DL}_d) \qquad \max_{\mu \in \Lambda(x_0, u_0)} D_u L(x_0, \mu, u_0) d, \tag{5.323}$$

其中

$$L(x,\mu,u) := f(x,u) + \int_{\Omega} g(x,u,\omega)\mathrm{d}\mu(\omega)$$

是 Lagrange 函数, $\Lambda(x_0,u_0)$ 是满足一阶最优性条件 (5.278) 的 Lagrange 乘子 $\mu \in C(\Omega)^*$ 的集合.

注 5.120　注意到, 即使 $\bar{x}(u)$ 是连续的, $\mu(u)$ 是唯一的, 对应于 $\bar{x}(u)$ 的 Lagrange 乘子 $\mu(u) \in C(\Omega)^*$ 也可能是 u 的非连续函数 (关于 $C(\Omega)^*$ 的全变差范数). 例如, 设 $\mu(u) = \lambda(u)\delta(\omega(u))$, 其中 $u \to \lambda(u) \in \mathbb{R}, u \to \omega(u) \in \Omega$. 假如 $u \neq u', \|\mu(u) - \mu(u')\| = |\lambda(u)| + |\lambda(u')|$. 若 $\lambda(u_0) \neq 0$ 且 $\omega(u)$ 在 u_0 的邻域内不是常数, 则 $\mu(u)$ 在点 u_0 处是非连续的.

这时方向正则性条件 (见 4.2 节) 可以表述为下述形式: 存在 $h \in \mathbb{R}^n$ 满足

$$h \cdot \nabla_x g_\omega(x_0,u_0) + D_u g_\omega(x_0,u_0)d < 0, \quad \forall \omega \in \Delta(x_0,u_0). \tag{5.324}$$

若广义 Mangasarian-Fromovitz 约束规范(5.279) 成立, 则对任何方向 d, 上述条件 (5.324) 是成立的. 条件 (5.324) 是线性化问题 (PL_d) 的 Slater 条件, 所以 (与命题 4.21 相比较), 若条件 (5.324) 成立, 则 $\mathrm{val}(\mathrm{PL}_d)=\mathrm{val}(\mathrm{DL}_d)< \infty$, 且 (PL_d) 与 (DL_d) 共同的最优值是有限的充分必要条件是 $\Lambda(x_0,u_0)$ 是非空的. 此种情况, (DL_d) 的最优解集合 $\mathcal{S}(\mathrm{DL}_d)$ 是 $\Lambda(x_0,u_0)$ 的非空的有界的子集. 由命题 5.106 可得, 若条件 (5.324) 成立, 则对偶问题(DL_d) 中的极大化运算可以仅仅关于离散测度进行. 即对任何 $m \geqslant n+1$, (DL_d) 的最优值与问题:

$$\max_{\mu \in \Lambda_m(x_0,u_0)} D_u L(x_0,\mu,u_0)d \tag{5.325}$$

的最优值相同, 其中 $\Lambda_m(x_0,u_0)$ 是满足一阶条件 (5.274) 的离散测度 $\mu = \sum_{i=1}^{m} \lambda_i \delta(\omega_i)$ 的集合. 注意到, 对离散测度 $\mu = \sum_{i=1}^{m} \lambda_i \delta(\omega_i)$, 有

$$D_u L(x_0,\mu,u_0)d = D_u f(x_0,u_0)d + \sum_{i=1}^{m} \lambda_i D_u g(x_0,u_0,\omega_i)d.$$

结合上述讨论与命题 4.22, 得到下述结果.

命题 5.121　设 x_0 是 (P) 的最优解, 方向正则性条件 (5.324) 成立, 则问题 (PL_d), (DL_d), $m = n+1$ 时的问题 (5.325) 的最优值是相同的, 且

$$\limsup_{\substack{t \downarrow 0 \\ d' \to d}} \frac{v(u_0 + td') - v(u_0)}{t} \leqslant \mathrm{val}(\mathrm{DL}_d). \tag{5.326}$$

设非扰动问题 (P) 的最优解集 $\mathcal{S}(u_0)$ 是非空的且在每一点 $x_0 \in \mathcal{S}(u_0)$ 处方向正则性条件成立, 则由 (5.326) 可得

$$\limsup_{\substack{t \downarrow 0 \\ d' \to d}} \frac{v(u_0 + td') - v(u_0)}{t} \leqslant \inf_{x \in \mathcal{S}(u_0)} \sup_{\mu \in \Lambda(x, u_0)} D_u L(x_0, \mu, u_0) d. \tag{5.327}$$

上述不等式的右端的 Lagrange 乘子测度集合 $\Lambda(x, u_0)$ 可由 $m = n + 1$ 时的离散测度子集 $\Lambda_m(x, u_0)$ 代替. 有趣的是, 上述不等式可推出, 若对 (P) 的至少一最优解 x_0, 集合 $\Lambda(x_0, u_0)$ 是空集且方向正则性条件成立, 则 $v'(u_0, d) = -\infty$. 在这样的情况下, 还是可以推导最优值函数变化的 $t^{1/2}$ 阶的上方估计和下方估计 (见 4.8.3 节).

如例 4.23, 即使 (P) 具有唯一的最优解, 由 (5.327) 右端给出的上界也可能不是精确的, 即在某些情况下, $v(u)$ 的一阶方向导数公式涉及 (P_u) 的二阶信息 (见 4.8 节). 在 4.3.2 节中已经证得, 若 (P) 的最优解集 $\mathcal{S}(u_0)$ 是非空的, 在 $\mathcal{S}(u_0)$ 的每一点处方向正则性条件均成立, 下确界–紧致性条件成立且下述的条件之一成立: (i) 问题 (P) 是凸的, (ii)(P_u) 具有 Lipschitz 稳定的最优解, (iii) 对每一 $x \in \mathcal{S}(u_0), \Lambda(x, u_0)$ 是单点集 (对确保 Lagrange 乘子唯一性的条件, 见定理 5.114), 则 $v(u)$ 在 u_0 处是方向可微的, 且 $v'(u_0, d)$ 由 (5.327) 的右端给出.

现在讨论 (P_u) 的最优解的定量稳定性. 这一节的余下部分, 假设函数 $f(x, u)$, $g_\omega(x, u)$ 是二次连续可微的且 $D^2 g_\omega(x, u)$ 在 $\mathbb{R}^n \times U \times \Omega$ 上是连续的, 可推出, 从 $\mathbb{R}^n \times U$ 到 $C(\Omega)$ 的对应的映射 $G(x, u)(\cdot) := g(x, u, \cdot)$ 是二次连续可微的. 考虑方向 $d \in U$, 当 $u \to u_0$ 时, 最优解 $\bar{x}(u) \in \mathcal{S}(u)$ 收敛到 $x_0 \in \mathcal{S}(u_0)$ 称为是方向 Lipschitz 稳定的, 若对 $t \geqslant 0$, $\|\bar{x}(u_0 + td) - x_0\| = O(t)$. 二阶充分性条件的强形式是在相应的二阶充分条件中将集合 $\Lambda(x_0, u_0)$ 用它的子集即 (DL_d) 的最优解集 $\mathcal{S}(DL_d)$ 代替得到的.

根据定理 4.55, 若下述条件成立, 则 $\bar{x}(u)$ 是方向 Lipschitz 稳定的: (i) 集合 $\Lambda(x_0, u_0)$ 是非空的, (ii) 方向正则性条件 (5.324) 成立, (iii) 二阶充分性条件的强形式成立, (iv) 线性化问题 (PL_d) 有最优解 $\hat{h} = \hat{h}(d)$, 满足

$$\sup_{\omega \in \Omega} \left\{ g_\omega(x_0, u_0) + t D g_\omega(x_0, u_0)(\hat{h}, d) \right\} \leqslant \kappa t^2 \tag{5.328}$$

对所有充分小的 $t \geqslant 0$ 成立, 其中 $\kappa > 0$ 是常数. 注意到条件 (5.328) 是定理 4.55 用到的抽象条件 (4.142) 对当前的半无限规划情况的具体形式.

相反地, $\Lambda(x_0, u_0)$ 的非空性质与 (PL_d) 的最优解的存在性对 $\bar{x}(u)$ 的上述方向 Lipschitz 稳定性是必要性条件. 强二阶充分性条件对于 $\bar{x}(u)$ 的方向 Lipschitz 稳定性是 “几乎必要的”(见定理 4.91 证明后面的注记). 例 4.63 表明, 条件 (5.328) 对于 $\bar{x}(u)$ 的方向 Lipschitz 稳定性也是本质性的.

最优化问题 (PL_d) 是线性半无限规划问题. 可能发生即使 (PL_d) 的最优值是有限的, 也可能没有最优解的情况 (见下述例子). 此种情况下, $\bar{x}(u)$ 沿方向 d 不是 Lipschitz 稳定的. 若 $\Delta(x_0,u_0)$ 是有限的, 则 (PL_d) 成为线性规划问题, 因此, 若它的最优值是有限的, 则它具有最优解.

若 (PL_d) 具有最优解 $\hat{h}$, 则应该满足 (PL_d) 的约束的可行性. 因为对所有的 $\omega \in \Delta(x_0,u_0), g_\omega(x_0,u_0)=0$, 由于这些约束的可行性, 如果 $\Delta(x_0,u_0)=\Omega$, 则条件 (5.328) 是自动成立的. 现在设 Ω 是赋范线性空间的紧致子集, $D_{x,u}g(x_0,u_0,\cdot)$ 在 Ω 上是 Lipschitz 连续的. 令 $\overline{\omega}(t)$ 是 (5.328) 左端括弧内的函数在 Ω 上的最大值点, 如果对 $t \geqslant 0, \mathrm{dist}(\overline{\omega}(t),\Delta(x_0,u_0))=O(t)$, 则条件 (5.328) 成立. 由命题 4.32, 这一条件可由下述二阶增长条件确保：存在 $c>0$ 与 $\Delta(x_0,u_0)$ 的邻域 $\mathcal{N}$, 满足

$$-g(x_0,u_0,\omega) \geqslant c[\mathrm{dist}(\omega,\Delta(x_0,u_0))]^2, \quad \forall \omega \in \Omega \cap \mathcal{N}.$$

这是下层问题的二阶增长条件, 即条件 (5.328) 可由上述下层问题的二阶增长条件推出.

还注意到, 若方向正则性条件 (5.324) 成立, 则 (PL_d) 与 (DL_d) 间不存在对偶间隙, 且若 $\mu=\sum_{i=1}^{m}\lambda_i\delta(\omega_i)$ 是 (DL_d) 的最优解, 则 (PL_d) 的最优解集 $\mathcal{S}(\mathrm{PL}_d)$ 可以表示为如下形式：

$$\left\{h : \sum_{i=1}^{m}\lambda_i Dg_{\omega_i}(x_0,u_0)(h,d)=0,\ Dg_\omega(x_0,u_0)(h,d)\leqslant 0,\ \omega\in\Delta(x_0,u_0)\right\}.$$

这可以通过将 (4.50) 给出的 $\mathcal{S}(\mathrm{PL}_d)$ 的抽象公式用到当前的情况而得到.

例 5.122　考虑半无限规划问题：

$$\min_{x\in\mathbb{R}^2} x_1+x_1^2+x_2^2 \quad \text{s. t. } -\omega x_1-(1-\omega)x_2+u(\omega)\leqslant 0, \quad \omega\in[0,1], \tag{5.329}$$

其参数为 $u \in C[0,1]$. 对于 $u_0 \equiv 0$, 相应的 (非扰动) 问题的可行域是 $\mathbb{R}^2_+$ 且 $x_0=(0,0)$ 是唯一的最优解. Slater 条件成立, 且 (P_u) 的可行集对任何 $u\in C[0,1]$ 是非空的, 问题 (P_u) 是凸的, 其目标函数是强凸的, 因此, 它有唯一最优解 $\bar{x}(u)$. 由于下确界–紧致性条件成立, 有 $\bar{x}(\cdot)$ 在 $C[0,1]$ 上是连续的.

二阶充分条件的强形式在最优解 x_0 处是成立的. 这里有非扰动问题的所有约束在解 x_0 处均是起作用的, 因此 $\Delta(x_0,u_0)=[0,1]$. 所以, 对给定的方向 $d\in C[0,1]$, 线性化问题是

$$\min_{h\in\mathbb{R}^2} h_1 \quad \text{s. t.} -\omega h_1-(1-\omega)h_2+d(\omega)\leqslant 0, \quad \omega\in[0,1]. \tag{5.330}$$

令 $d(\cdot)$ 满足, 对所有的 $\omega \in (0,1)$ 有 $d(\omega) > 0, d(0) = d(1) = 0$, 则问题 (5.330) 的可行集被包含在集 $\mathbb{R}^2_+$ 中, 这一可行集与轴 h_1 有公共点的充分必要条件是

$$\limsup_{\omega \downarrow 0} \frac{\mathrm{d}(\omega)}{\omega} < \infty. \tag{5.331}$$

如果 (5.331) 左端的极限是 ∞ 的, 则 (5.330) 的可行集渐近地接近轴 h_1. 所以, 线性化问题 (5.330) 有最优解的充分必要条件是条件 (5.331) 成立. 注意到, 由于 $\Delta(x_0, u_0) = [0,1]$, 在线性化问题 (5.330) 有最优解的前提下, 条件 (5.328) 对任何 d, 是成立的, 得到 $\bar{x}(u)$ 在 $u \equiv 0$ 处沿方向 d Lipschitz 稳定的充分必要条件是条件 (5.331) 成立.

注意到, 在上述例子中, 最优解 $\bar{x}(u)$ 沿某些方向是方向 Lipschitz 稳定的, 而沿其他方向则不是方向 Lipschitz 稳定的. 为了确保一致 Lipschitz 稳定性$\|\bar{x}(u) - x_0\| = O(\|u - u_0\|)$, 需要更强的条件. 尤其, 线性化问题($\mathrm{PL}_d$) 应具有最优解 $\hat{h}(d)$ 满足 $\|\hat{h}(d)\| = \max O(\|d\|)$(见定理 4.65 与这一定理后面的讨论).

定理 4.101 给出了关于最优值函数的二阶展开与相应的最优解的方向可微性的抽象形式的基本灵敏度结果. 对下述形式的路径:

$$u(t) := u_0 + td + \frac{1}{2}t^2 r + o(t^2), \quad t \geqslant 0,$$

相应的极小–极大问题是

$$(\mathcal{DQ}) \qquad \min_{h \in \mathcal{S}(\mathrm{PL}_d)} \max_{\mu \in \mathcal{S}(\mathrm{DL}_d)} \xi(h, \mu), \tag{5.332}$$

其中

$$\xi(h,\mu) := D_u L(x_0, \mu, u_0) r + D^2 L(x_0, \mu, u_0)((h,d),(h,d)) - \sigma(\mu, T_k^{i,2}(y,z)),$$

其中 $K := C_-(\Omega)$, $y(\cdot) := g(x_0, u_0, \cdot)$, $z(\cdot) := D_{x,u} g(x_0, u_0, \cdot)(h, d)$.

为应用定理 4.101, 需要计算 $\xi(h,\mu)$ 的上述定义中的 sigma 项, 并且验证 $C_-(\Omega)$ 在点 y 处沿方向 z 关于线性映射 $D_x G(x_0, u_0)$ 的二阶正则性. 由前一节最后的讨论, 存在函数 $\tau_{y,z} : \Omega \to \mathbb{R}$ 满足, 对 $\mu = \sum_{i=1}^m \lambda_i \delta(\omega_i)$,

$$\sigma(\mu, T_K^{i,2}(y,z)) = \sum_{i=1}^m \lambda_i \tau_{y,z}(\omega_i).$$

进一步, 如定理 5.118 所示, 有 $C_-(\Omega)$ 在点 y 处沿方向 z 是二阶正则的, 且 $\tau_{y,z}(\cdot)$ 由问题 (5.300) 的最优值给出, 即如果 (i) 集合 $\Delta(x_0, u_0)$ 是有限的, (ii) 集合 Ω 在 $\Delta(x_0, u_0)$ 的每一点均是二阶正则的, (iii) 下层问题的二阶增长条件是成立的, 则方程 (5.320) 是成立的.

第6章　最优控制

6.1　引　　言

这一章给出抽象优化理论在某些偏微分方程的最优控制问题的应用. 这些素材自成体系, 仅需要基本积分理论的某些知识.

这一章如下组织：6.2 节回顾线性与半线性椭圆方程以及与之相联系的泛函空间理论的某些基本结果. 6.3 节讨论半线性椭圆方程的最优控制. 目标函数是二次的, 问题加在控制上的约束在定义 3.5.1 意义下是多面性的. 对大于 3 的空间维数, 对控制空间用两个范数, 即 $L_2(\Omega)$ 与 $L_s(\Omega)$, 其中 $s > \dfrac{1}{2}n$ 是正数. 在多面性的控制约束的情形给出完整的二阶理论及扰动分析. 在积分状态约束与分布状态约束的情形给出部分结果. 还讨论了一种非单调非线性的情况, 此种情况状态方程不是适定的, 我们得到非规范形式的最优性系统. 6.4 节讨论最简单变分不等式, 即障碍问题的最优控制. 因为障碍问题的解是优化问题的解, 为推导一阶灵敏度分析, 可以应用前面得到的某些结果.

6.2　线性与半线性椭圆方程

这一节给出后面将用到的线性与半线性椭圆方程理论的某些基本结果.

6.2.1　Dirichlet 问题

设 Ω 是 $\mathbb{R}^n$ 的有界开子集, 它具有 C^2 光滑边界 $\partial\Omega$, 生成元素记为 ω. 这一节的目的在于 Dirichlet 问题：

$$\begin{cases} -\Delta y = f, & 在\ \Omega\ 内, \\ y = 0, & 在\ \partial\Omega\ 上 \end{cases} \tag{6.1}$$

的研究, 这里 Δ 是 Laplace 算子：

$$\Delta y := \sum_{i=1}^{n} \frac{\partial^2 y}{\partial \omega_i^2},$$

f 是给定的函数或分布, y 是未知数. (6.1) 的第一个方程是 Poisson 方程, 第二个方程是齐次的 Dirichlet 边界条件.

这一问题的分析涉及几个泛函空间. 因此, 需要建立泛函空间嵌入的某些一般性质.

定义 6.1 设 V 与 H 是两个拓扑向量空间, 称 V 连续地嵌入到 H, 若存在连续的线性的一对一映射 $I: V \to H$. 称嵌入是稠密的, 若 $I(V)$ 是 H 的稠密子集.

因为本书所讨论的所有嵌入均是连续的嵌入, "嵌入" 在这里理解为 "连续的嵌入". 还经常用 v 代替 I_v, 即将 V 与 $I(V)$ 等同, 且写为 $V \subset H$. 若 V 与 H 是赋范空间, V 被嵌入到 H, 则存在 $c > 0$,

$$\|v\|_H \leqslant c\|v\|_V, \quad 对所有的\ v \in V.$$

引理 6.2 设 V 与 H 是两个局部凸的拓扑向量空间, $I: V \to H$ 是稠密嵌入. 则在下述意义下 H^* 被连续地嵌入到 V^*: 由下式定义的伴随映射 $I^* \to V^*$,

$$\langle I^*h^*, v\rangle_{V^*,V} := \langle h^*, v\rangle_{H^*,H}, \quad 对所有的\ h^* \in H,\ v \in V \tag{6.2}$$

是线性的连续的一一对应的.

证明 因为 I^* 是由 V 到 H 的内射的伴随算子, 且是线性的连续的. 现在证 I^* 是一一对应的. 若 $I^*h^* = 0$, 则由于 V 是 H 的稠密子集, 由 (6.2) 可得, 对所有的 $h \in H$, $\langle h^*,\ h\rangle = 0$, 从而 $h^* = 0$. □

若 H 与 V 满足上述引理的假设, 且 H 是 Hilbert 空间, 则由 Riesz 表现定理2.3.4, 可以将 H 与 H^* 等同, 因此有下述的包含关系:

$$V \subset H \subset V^*. \tag{6.3}$$

引理 6.3 令 H 是 Hilbert 空间, 与它的对偶空间相同, 令 V 是自反的Banach 空间, 稠密地嵌入到 H, 则 H(关于强拓扑) 是 V^* 的稠密子集.

证明 假设 H 不是 V^* 的稠密子集, 则由第二分离定理 2.14, 存在 V^* 上的非零的连续的线性形式 l 满足它在 H^* 取零值. 由于是自反的, V 与它的双重对偶是同构的. 因此, 存在 $v \in V$ 满足 $l(v^*) = \langle v^*,\ v\rangle_{V,\ V^*}$ 对所有的 $v^* \in V^*$ 均成立. 由 (6.2) 得到, 对所有的 $h \in H$, $\langle h,\ v\rangle_H = 0$, 取 $h = v$, 得到 $v = 0$, 这与假设矛盾. □

将上述结果用于 $H = L_2(\Omega)$, 在后续的讨论中, 总是把它与其对偶等同起来.

令 $q = (q_1, \cdots, q_l)$ 是非负整数分量的向量, 称为多重指标. 定义

$$|q| := q_1 + \cdots + q_l; \quad D^q y := \frac{\partial^{|q|} y}{\partial x_1^{q_1} \cdots \partial x_l^{q_l}}; \quad D^0 f = f.$$

回顾分布意义下的导数的概念. 设 $\mathcal{D}(\Omega)$ 是 Ω 上的, 其支撑是 Ω 的紧致的实值的 C^∞ 光滑函数的集合. 赋予 $\mathcal{D}(\Omega)$ 以下述的拓扑. 给定 Ω 的紧致子集 K, 令 $\mathcal{D}_K(\Omega)$

是定义在 Ω 上支撑包含在 K 中的实值 C^∞ 光滑函数的全体. 显然, $\mathcal{D}(\Omega)$ 是所有可能 K 的 $\mathcal{D}_K(\Omega)$ 的并. 赋予 $\mathcal{D}_K(\Omega)$ 以通过下述桶集族:

$$\mathcal{O}_{K,m} := \{g \in \mathcal{D}_K(\Omega) : \|D^q g\|_\infty \leqslant 1, \|q\| \leqslant m\}$$

引导的拓扑, 其中 m 是非负整数. 很清楚, $\mathcal{D}_K(\Omega)$ 是局部凸的可分的拓扑向量空间. 定义 $\mathcal{D}(\Omega)$ 的开集基本系统为集合 $\mathcal{O} \subset \mathcal{D}(\Omega)$ 的平移 (见 2.1.1 节), 其中 $\mathcal{O}$ 是桶集, 且对于每一紧致集 K, $\mathcal{O} \cap \mathcal{D}_K(\Omega)$ 是 $\mathcal{D}_K(\Omega)$ 的开子集. $\mathcal{D}(\Omega)$ 的子集 $\mathcal{G}$ 称为是开的, 若 $\mathcal{G}$ 的每一点均包含在含于 $\mathcal{G}$ 的某一基本开集中. 可以验证, $\mathcal{D}(\Omega)$ 是局部凸的可分的拓扑向量空间, 称为空间 $\mathcal{D}_K(\Omega)$ 的引导极限, 其中 K(像以前那样) 是 Ω 的紧致子集, 则在 $\mathcal{D}(\Omega)$ 中 $f_n \to f$ 当且仅当下述两个条件成立: (i) 存在 Ω 的紧致子集 K 满足每一 f_n 的支撑都包含于 K 中, 且 (ii) 对每一多重指标 q, $D^q f_n \to D^q f$, 收敛是一致的.

$\mathcal{D}(\Omega)$ 的拓扑对偶空间, 记为 $\mathcal{D}'(\Omega)$, 称为在 Ω 上的分布的空间. 所以, 若在 $\mathcal{D}(\Omega)$ 中 $f_n \to f$, 则对所有的 $f^* \in \mathcal{D}'(\Omega)$, 均有 $\langle f^*, f_n\rangle \to \langle f^*, f\rangle$.

给定分布f^* 与多重指标 q, 定义 $D^q f^*$ 如下:

$$\langle D^q f^*, z\rangle = (-1)^{|q|}\langle f^*, D^q z\rangle, \quad \forall z \in \mathcal{D}(\Omega).$$

微分算子 D^q 是 $\mathcal{D}'(\Omega)$ 中线性的连续映射. 令 $f^* \in \mathcal{D}'(\Omega)$. 称 $f^* \in L_1(\Omega)$, 若存在 $g \in L_1(\Omega)$, 满足

$$\langle f^*, f\rangle = \int_\Omega g(\omega) f(\omega)\mathrm{d}\omega, \quad \forall f \in \mathcal{D}(\Omega).$$

相反地, 给定 $g \in L_1(\Omega)$, 上述等式定义了一个分布, 把它等同于 g. 所以, 可以谈分布意义下的可积函数的导数. 若 $f \in L_1(\Omega)$ 是连续的 $|q|$ 阶可微的, 则 $D^q f$ 理解为分布意义下的导数, 与通常意义下的导数是重合的. 在 $|q| = 1$ 的情况, 这是 "分部积分" 公式的一个结果, 一般的情况可由归纳法去证明.

用 $\|\cdot\|_s$ 记 $L_s(\Omega)$ 的范数, $L_2(\Omega)$ 的数积记为 $\langle\cdot,\cdot\rangle_2$. 若 g 属于 $L_s(\Omega, \mathbb{R}^q)$, 其中 q 是正整数, 也记它的范数为 $\|g\|_s$. Sobolev 空间$W^{m,s}(\Omega)$, 其中 m 是非负的整数, $s \in [1, +\infty]$, 如下定义且赋予如下的范数:

$$\begin{aligned} W^{m,s}(\Omega) &:= \{y \in L_s(\Omega) : D^q y \in L_s(\Omega), \text{只要 } |q| \leqslant m\}, \\ \|y\|_{m,s} &:= \sum_{0 \leqslant |q| \leqslant m} \|D^q y\|_s. \end{aligned}$$

当 $s = 2$ 时, 用 $H^m(\Omega)$ 记 $W^{m,2}(\Omega)$, 赋予它如下的范数:

$$\|y\|_{m,2} := \left(\sum_{0 \leqslant |q| \leqslant m} \|D^q y\|_2^2\right)^{1/2},$$

空间 $W^{m,2}(\Omega)(H^m(\Omega))$, 赋予上述范数是 Banach(Hilbert) 空间. 最后, $W_0^{m,s}(\Omega)$ 记 $\mathcal{D}(\Omega)$ 在 $W^{m,s}(\Omega)$ 中的闭包, $H_0^m(\Omega) := W_0^{m,2}(\Omega)$. 令 $s' \in [1,\infty]$ 是 s 的共轭数, 即 $1/s + 1/s' = 1$. 令 $C(\overline{\Omega})(C_0(\Omega))$ 是 $\overline{\Omega}$ 上连续函数的集合 (满足在 $\partial\Omega$ 上取零值在 $\overline{\Omega}$ 上的连续函数的集合).

下述的 Sobolev 嵌入成立.

引理 6.4 令 Ω 是 $\mathbb{R}^n$ 的开子集, 具有 C^2 光滑边界, 则对 $m = 1, 2, s \in [1,\infty]$, 下述的连续包含成立:

$$\begin{cases} W^{m,s}(\Omega) \subset L_q(\Omega), & \dfrac{1}{q} \geqslant \dfrac{1}{s} - \dfrac{m}{n}, \quad 若\ \dfrac{1}{s} > \dfrac{m}{n}, \\ W^{m,s}(\Omega) \subset C(\overline{\Omega}), & 若\ \dfrac{1}{s} < \dfrac{m}{n}. \end{cases} \tag{6.4}$$

额外地, 第二个包含关系是紧的, 只要 $1/q > 1/s - m/n$, 第一包含关系也是紧的.

将经常用到 Sobolev 嵌入的下述结果:

$$H^1(\Omega) \overset{n=1}{\subset} C(\overline{\Omega}), \quad H^2(\Omega) \overset{n\leqslant 4}{\subset} C(\overline{\Omega}), \quad W^{2,s}(\Omega) \overset{s>n/2}{\subset} C(\overline{\Omega}), \tag{6.5}$$

其中, 如果包含符号上面的条件是成立的, 则包含关系是有效的.

若 $X_1 \subset X$, $A : X_1 \to Y$, 则称映射 $B : X \to Y$ 是 A 的扩展, 若 B 到 X_1 上的限制与 A 是重合的.

引理 6.5 (一致扩展原理) 设 X 与 Y 是两个 Banach 空间. 考虑线性映射 $A : X_1 \to Y$, 其中 X_1 是 X 的稠密的线性子空间, 存在 $c > 0$ 满足, 对所有的 $x_1 \in X_1, \|Ax_1\|_Y \leqslant c\|x_1\|_X$, 则 A 具有唯一的线性的到 X 的连续扩展.

证明 若 A 的两个连续的扩展在稠密的子集上是相等的, 则 (由连续性) 它们是处处相等的. 唯一性得证.

由 $\bar{A}x = \lim_k Ax_k$ 定义线性算子$\bar{A} : X \to Y$, 其中 $\{x_k\} \subset X_1$, 在 X 中 $x_k \to x$. 因为 $\{Ax_k\}$ 是 Y 中的 Cauchy 序列, 极限是存在的, 如果 $\{x_k'\} \subset X_1$, 在 X 中 $x_k' \to x$, 则 $\|A(x_k' - x_k)\| \leqslant c\|x_k' - x_k\| \to 0$, 从而这一极限不依赖于特取的序列. $\bar{A}$ 的线性性由定义可以得到, 由关系 $\|\bar{A}x\|_Y \leqslant c\|x\|_X$ 可推出 $\bar{A}$ 的连续性. □

后面经常将相对于 ω_i 的偏导数用记号 $\partial\cdot/\partial\omega_i$ 表示. 给定 $y \in \mathcal{D}(\Omega)$, 用 $\nabla y(\omega)$ 记 $\mathcal{D}(\Omega)^n$ 中的元素, 它的第 i 个分量是 $\partial y/\partial\omega_i$.

引理 6.6 (Green 公式) 令 $s \in [1,\infty), s'$ 是它的共轭数. 则下述结论成立:

(i) 若 $f \in L_{s'}(\Omega)$, 则分布 $\dfrac{\partial f}{\partial\omega_i}$ 在 $W_0^{1,s}(\Omega)$ 上有连续的扩展, 满足

$$\left\langle \frac{\partial f}{\partial\omega_i}, y \right\rangle = -\int_\Omega f(\omega)\frac{\partial y_i}{\partial\omega_i}(\omega)\mathrm{d}\omega, \quad \forall y \in W_0^{1,s}(\Omega). \tag{6.6}$$

(ii) 令 y 属于 $W^{1,s'}(\Omega)$, 则分布 Δy 在 $W_0^{1,s}(\Omega)$ 上有唯一的扩展, 且对每一 $z \in W_0^{1,s}(\Omega)$, 下述的Green公式成立:

$$-\langle \Delta y, z\rangle = \int_\Omega \nabla z(\omega) \cdot \nabla y(\omega) \mathrm{d}\omega. \tag{6.7}$$

证明 (i) 若 y 属于 $\mathcal{D}(\Omega)$, 则 (6.6) 中的等式成立且

$$\left|\left\langle \frac{\partial f}{\partial \omega_i}, y\right\rangle\right| \leqslant \|f\|_{s'}\|y\|_{1,s}.$$

由引理 6.5, $\mathcal{D}(\Omega)$ 在 $W_0^{1,s}(\Omega)$ 的稠密性可推出 $\left\langle \frac{\partial f}{\partial \omega_i}, y\right\rangle$ 在 $W_0^{1,s}(\Omega)$ 上有唯一的扩展, 满足 (6.6).

(ii) 只要 z 属于 $\mathcal{D}(\Omega)$, Green 公式(6.7) 是成立的, 此种情况,

$$|\langle \Delta y, z\rangle| = \left|\int_\Omega \nabla z(\omega)\nabla y(\omega)\mathrm{d}\omega\right| \leqslant \|z\|_{1,s}\|y\|_{1,s'}.$$

所以, $z \to \langle \Delta y, z\rangle$ 相对于 $W_0^{1,s}(\Omega)$ 的拓扑是连续的. 由引理 6.5, $\mathcal{D}(\Omega) \subset W_0^{1,s}(\Omega)$ 的稠密性推出, 当 $z \in W_0^{1,s}(\Omega)$ 时, 它有唯一的扩展. □

不引用记号, 有时把对偶性乘积写为积分的形式. 例如, 若 $y \in W_0^{1,s'}(\Omega)$, $z \in W_0^{1,s}(\Omega)$, 将 $\int_\Omega \Delta y(\omega) z(\omega)\mathrm{d}\omega$ 理解为相应的对偶性乘积.

$W_0^{1,s}(\Omega)$ 的对偶空间 ($H_0^1(\Omega)$ 的对偶空间) 记为 $W_0^{-1,s'}(\Omega)(H^{-1}(\Omega))$. 这一记号可通过下述引理得到检验.

引理 6.7 令 $s \in [1,\infty)$, 则每一 $f^* \in W_0^{-1,s'}(\Omega)$ 具有下述形式:

$$f^* = f_0 + \sum_{i=1}^n \frac{\partial f_i}{\partial \omega_i}, \tag{6.8}$$

其中 $f_i, i = 0, \cdots, n$, 属于 $L_{s'}(\Omega)$.

证明 考虑映射 $T: y \to (y, \nabla y)$, 其定义域为 $W_0^{1,s}(\Omega)$, 其像在 $Y := L_s(\Omega) \times (L_s(\Omega))^n$ 中. 显然, T 是线性的连续的一一对应的, 甚至是等距的, 即对所有的 $y \in W_0^{1,s}(\Omega)$, 有 $\|T(y)\| = \|y\|_{1,s}$. 所以, 它的值域 $\mathcal{R}(T)$ 是闭的且 T^{-1}: $\mathcal{R}(T) \to W_0^{1,s}(\Omega)$ 是连续的算子. 令 $f^* \in W_0^{-1,s}(\Omega)$, 则算子$f^* \circ T^{-1}$ 属于 $\mathcal{R}(T)$ 的拓扑对偶. 由 Hahn-Banach 定理2.10, 存在 Y^* 上的 $f^* \circ T^{-1}$ 的线性的连续的扩展. 因为 $Y^* = L_{s'}(\Omega) \times (L_{s'}(\Omega))^n$, 则存在 $L_{s'}(\Omega)$ 中的 $f_0, \cdots, f_n$, 满足对所有的 $y \in W_0^{1,s}(\Omega)$,

$$\langle f^*, y\rangle = \int_\Omega f_0(\omega) y(\omega)\mathrm{d}\omega - \sum_{i=1}^n \int_\Omega f_i(\omega) \frac{\partial y_i}{\partial \omega_i}(\omega)\mathrm{d}\omega.$$

由引理 6.6(i), f^* 具有该引理所述的形式. 相反地, 由 (6.6), 具有形式 (6.8) 的任何分布均定义了 $W_0^{1,s}(\Omega)$ 上的连续的线性形式. 结果得证. □

由上述结果可推出, 对 $s \in [2,\infty)$, 有下述的连续包含链:

$$\mathcal{D}(\Omega) \subset W_0^{1,s}(\Omega) \subset L_2(\Omega) \subset W_0^{-1,s'}(\Omega) \subset \mathcal{D}'(\Omega).$$

注意到, 前三个包含关系是稠密的.

引理 6.8 (Poincaré 不等式) *令 $s \in [1,\infty]$, 若 Ω 是有界的, 且存在 $c > 0$ 满足, 对所有的 $f \in W_0^{1,s}(\Omega)$, 有*

$$\|f\|_s \leqslant c\|\nabla f\|_s. \tag{6.9}$$

证明 因为 $s = \infty$ 的情况很容易证明, 我们给出 $s \in [1,\infty)$ 时的证明. 因为 $f \mapsto (\|f\|_s, \|\nabla f\|_s)$ 是 $W_0^{1,s}(\Omega)$ 上的连续函数, 只需证明当 f 属于 $\mathcal{D}(\Omega)$ 的稠密子集时 (6.9) 成立. 令 a, b 满足当 $\omega \in \Omega$ 时 $\omega_1 \in (a,b)$. 置 $\omega = (\omega_1, \widetilde{\omega})$. 令 s' 是 s 的共轭数, e_1 是 $\mathbb{R}^n$ 的自然基的第一个元素. 将 $f \in \mathcal{D}(\Omega)$ 扩展为在 Ω 外面取 0 值, 用 Hölder 不等式, 得到

$$|f(\omega)| = \left|\int_a^{\omega_1} \frac{\partial f}{\partial \omega_1}(\sigma, \widetilde{\omega})\mathrm{d}\sigma\right| \leqslant (b-a)^{1/s'} \left(\int_a^b \left|\frac{\partial f}{\partial \omega_1}(\sigma, \widetilde{\omega})\right|^s \mathrm{d}\sigma\right)^{1/s}.$$

由 Fubini 定理, 用到 $s = 1 + s/s'$, 得到 $\|f\|_s^s$ 等于

$$\begin{aligned}\int_{\mathbb{R}^{n-1}} \int_a^b |f(t,\widetilde{\omega})|^s \mathrm{d}t\mathrm{d}\widetilde{\omega} &\leqslant (b-a)^s \int_{\mathbb{R}^{n-1}} \int_a^b \left|\frac{\partial f}{\partial \omega_1}(\sigma, \widetilde{\omega})\right|^s \mathrm{d}\sigma\mathrm{d}\widetilde{\omega} \\ &\leqslant (b-a)^s \left\|\frac{\partial f}{\partial \omega_1}\right\|_s^s,\end{aligned}$$

因此, 若置 $c := b - a$, 则结论成立. □

由上述证明可得, 当 Ω 沿一方向是有界的, 则 Poincaré 不等式成立. 还注意到, 这一不等式使得 $c \leqslant \operatorname{diam}(\Omega)$, 其中 $\operatorname{diam}(\Omega)$ 是 Ω 的直径, 定义为包含 Ω 的球的半径的下确界①.

根据 Poincaré引理, 可以赋予 $W_0^{1,s}(\Omega)$ 以等价的范数:

$$\|y\|_{1,s,0} := \|\nabla y\|_s.$$

赋予对偶空间 $W^{-1,s'}(\Omega)$ 以相应的对偶范数, 定义为

$$\|\nabla f\|_{-1,s'} := \sup\{\langle f, y\rangle : \|y\|_{1,s,0} = 1\}.$$

① 应该是直径的下确界.

在 $s = s' = 2$ 的情况, $H_0^1(\Omega)$ 的范数与数积是相联系的:

$$\langle y, z\rangle = \int_\Omega \nabla y(\omega)\nabla z(\omega)\mathrm{d}\omega.$$

由 Riesz 表示定理2.34, 可推出下述结果.

引理 6.9 若 Ω 是有界的, 则方程

$$\int_\Omega \nabla y(\omega)\nabla z(\omega)\mathrm{d}\omega = \langle f, z\rangle, \quad 对所有的 z \in H_0^1(\Omega) \tag{6.10}$$

在 $H_0^1(\Omega)$ 中具有唯一的解, 其中 $f \in H^{-1}(\Omega)$.

将这一结论与引理 6.6 相结合得到, 若将 Poisson 方程 (6.1) 按分布意义理解, 将 Dirichlet 条件理解为属于 $H_0^1(\Omega)$ 的常数, 则下述结果成立.

引理 6.10 Dirichlet 问题 (6.1) 具有唯一解 $y \in H_0^1(\Omega)$, 且 $\|y\|_{1,2,0} = \|f\|_{-1,2}$, 其中 $f \in H^{-1}(\Omega)$.

方程 (6.10) 为 Dirichlet 问题的变分表示. 由引理 6.7 与 6.10, 得到 $H^{-1}(\Omega) = \Delta(H_0^1(\Omega))$.

Dirichlet 问题的解可以通过下述方式得到. 给定 $f \in H^{-1}(\Omega)$, 考虑如下定义的目标函数 $F : H_0^1(\Omega) \to \mathbb{R}$,

$$F(y) := \int_\Omega |\nabla y(\omega)|^2 \mathrm{d}\omega - \langle f, y\rangle.$$

根据 Poincaré不等式, 这是 $H_0^1(\Omega)$ 中强凸的且连续的二次形式, 由 (6.10) 刻画, 它在唯一点处取到最小. 这就是著名的 Dirichlet 原理. 由于这一结论是用于证明 Riesz 定理的具体情况, 两个证明本质上是相同的. 然而, 如将要看到的, 后者可被推广到研究某些半线性的椭圆方程.

6.2.2 半线性的椭圆方程

本章的余下部分, ϕ 是模 L 的 Lipschitz 连续的连续可微的, 记为 $\mathbb{R}$ 上定义的非递减的实值函数. 在某些叙述中, 设 ϕ 是二次连续可微的. 考虑半线性椭圆方程:

$$\begin{cases} -\Delta y + \phi(y) = u, & 在\ \Omega\ 内, \\ y = 0, & 在\ \partial\Omega\ 上, \end{cases} \tag{6.11}$$

这里 $\phi(y)$ 是几乎处处定义于 Ω 上的函数:

$$\phi(y)(\omega) := \phi(y(\omega)), \quad \text{a.e.} \ \ \omega \in \Omega.$$

不失一般性, 可设 $\phi(0) = 0$. 注意到, 在控制理论中传统的记号, u 表示控制参数, 不像本书其他章节, u 表示扰动参数.

引理 6.11 令 $s \in [1, \infty)$. 若 $u \in W_0^{1,s}(\Omega)$, 则 $\phi(u) \in W_0^{1,s}(\Omega)$, 满足 $\nabla\phi(u) = \phi'(u)\nabla u$ 在 Ω 上几乎处处成立.

证明 由 $W_0^{1,s}(\Omega)$ 的定义, 在 $\mathcal{D}(\Omega)$ 中存在序列 u_n 满足, 在 $W_0^{1,s}(\Omega)$ 中 $u_n \to u$. 很清楚, $\phi(u_n) \in W_0^{1,s}(\Omega)$ 且 $\nabla\phi(u_n) = \phi'(u_n)\nabla u_n$. 则几乎处处地,

$$\begin{aligned}|\nabla\phi(u_n) - \phi'(u)\nabla u| &\leqslant |\phi'(u_n) - \phi'(u)| \cdot |\nabla u| + |\phi'(u_n)| \cdot |\nabla u_n - \nabla u| \\ &\leqslant |\phi'(u_n) - \phi'(u)| \cdot |\nabla u| + L|\nabla u_n - \nabla u|.\end{aligned}$$

函数 $|\phi'(u_n) - \phi'(u)| \cdot |\nabla u|$ 被 $2L|\nabla u| \in L_s(\Omega)$ 控制, 几乎处处收敛到 0. 根据 Lebesgue 控制收敛定理, 在 $L_s(\Omega)$ 中, $|\phi'(u_n) - \phi'(u)| \cdot |\nabla u| \to 0$. 与上述不等式相联系, 得到 $\nabla\phi(u_n) \to \phi'(u)\nabla u$ 于 $L_s(\Omega)$ 中. 根据 Poincaré不等式, 在 $W_0^{1,s}(\Omega)$ 中 $\phi(u_n)$ 是 Cauchy 序列, 在 $L_s(\Omega)$ 中它的极限 v 满足 $v = \phi(u)$ 且 $\nabla v = \phi'(u)\nabla u$. 结论得证. □

命题 6.12 设 ϕ 是非递减的, 令 $u \in H^{-1}(\Omega)$, 则 (6.11) 具有唯一解 $y \in H_0^1(\Omega)$, 且 $\|y\|_{1,2,0} \leqslant \|u\|_{-1,2}$.

证明 令 Φ 是 ϕ 的不定积分, 满足 $\Phi(0) = 0$. 因为 ϕ 是非递减的, 非负的凸函数, Φ 满足 $|\Phi(t)| \leqslant \frac{1}{2}Lt^2$. 所以, 函数 $z \to \int_\Omega \Phi(z(\omega))\mathrm{d}\omega : L_2(\Omega) \to \mathbb{R}$ 是实值的凸的且上方有界的, 因而由命题 2.107 可得, 它是连续的, 可得下面的 Taylor 展开式

$$\int_\Omega \Phi(y + tz)(\omega)\mathrm{d}\omega = \int_\Omega \Phi(y)(\omega)\mathrm{d}\omega + t\int_\Omega \phi(y + \theta tz)(\omega)z\mathrm{d}\omega,$$

其中 $\theta \in (0, 1)$ 依赖于 ω. 因此, 对给定的 $z \in L_2(\Omega)$, 当 $t \downarrow 0$ 时, $\phi(y + \theta tz)(\omega)z \to \phi(y)z$ 在 Ω 内几乎处处成立, 且 $|\phi(y + \theta tz)| \leqslant |\phi(y)| + tL|z|$, 由 Lebesgue 定理可得 $\phi(y + \theta tz) \to \phi(y)$ 在 $L_2(\Omega)$ 内成立. 于是, 函数 $z \to \int_\Omega \Phi(z(\omega))\mathrm{d}\omega : H_0^1(\Omega) \to \mathbb{R}$ 具有 Gâteaux 导数 $\phi(y)$. 这一导数关于 y 是 Lipschitz 连续的, 具有相同的常数 L.

所以, 势函数

$$F(y) := \frac{1}{2}\int_\Omega |\nabla y(\omega)|^2\mathrm{d}\omega + \int_\Omega \Phi(y(\omega))\mathrm{d}\omega - \langle u, y\rangle$$

是连续可微函数. 作为强凸函数与凸函数的和, F 也是强凸的. 由引理 2.33, 它在唯一点 $y \in H_0^1(\Omega)$ 处取到极小. 根据引理 6.10 与上述的讨论, $DF(y) = -\Delta y + \phi(y) - u$. 所以, 方程 $DF(y) = 0$ 等价于 (6.11), 在 $H_0^1(\Omega)$ 中具有唯一的解. 最后, 计算关于 y 的 (6.11) 的对偶积, 得到

$$\int_\Omega |\nabla y(\omega)|^2\mathrm{d}\omega \leqslant \int_\Omega |\nabla y(\omega)|^2\mathrm{d}\omega + \int_\Omega \phi(y(\omega))y(\omega)\mathrm{d}\omega = \langle u, y\rangle \leqslant \|u\|_{-1,2}\|y\|_{1,2,0},$$

由此可得到估计式. □

命题 6.13　映射 $u \mapsto y_u$ ((6.11) 的解) 是从 $H^{-1}(\Omega)$ 到 $H_0^1(\Omega)$ 连续可微的, 它沿方向 v 的导数 z 是方程

$$\begin{cases} -\Delta z + \phi'(y_u)z = v, & \text{在 } \Omega \text{ 内}, \\ z = 0, & \text{在 } \partial\Omega \text{ 上} \end{cases} \tag{6.12}$$

的唯一解.

证明　(a) 首先证明 $u \mapsto y_u$ 是从 $H^{-1}(\Omega)$ 到 $H_0^1(\Omega)$ 的压缩映射. 令 u_1 与 u_2 属于 $H^{-1}(\Omega)$, 相联系的状态为 y_1 与 y_2. 置

$$v := u_2 - u_1, \quad z := y_2 - y_1$$

与

$$\psi(\omega) := \begin{cases} \dfrac{\phi(y_2)(\omega) - \phi(y_1)(\omega)}{(y_2 - y_1)(\omega)}, & \text{若 } y_2(\omega) \neq y_1(\omega), \\ \phi'(y_1)(\omega), & \text{否则}. \end{cases}$$

因为 ϕ 是 Lipschitz 连续的, 所以 $\|\psi\|_\infty \leqslant L$. 观察到 $z = 0$ 于 $\partial\Omega$ 上, 且

$$-\Delta z + \psi z = v. \tag{6.13}$$

对两边计算关于 z 的对偶积, 由 $\psi \geqslant 0$, 得到 $\|z\|_{1,2,0} \leqslant \|v\|_{-1,2}$.

(b) 在 $H^{-1}(\Omega)$ 中固定 u 与 v, 对 $t > 0$, 置

$$y = y_u, \quad u_t := u + tv, \quad y_t := y_{u_t}, \quad z_t := t^{-1}(y_t - y).$$

由步 (a) 的论证, 可得到 $-\Delta z_t + \psi_t z_t = v$, $\psi_t(\omega) \in \phi'([y(\omega), y_t(\omega)])$ 在 Ω 上几乎处处成立, 且 z_t 在 $H_0^1(\Omega)$ 内是一致有界的. 令 z 是 z_t 在 $H_0^1(\Omega)$ 内的弱极限点, 由引理 6.4, z 是 z_t 在 $L_2(\Omega)$ 内的强极限点. 因为 ψ_t 在 $L_\infty(\Omega)$ 内是有界的, 且在 Ω 上几乎处处收敛到 $\phi'(y)$, 由 Lebesgue 定理可得 $\psi_t z_t = \psi_t(z_t - z) + z\psi_t$ 在 $L_2(\Omega)$ 收敛到 $\phi'(y)z$. 所以, 在 $H^{-1}(\Omega)$ 中 $-\Delta z_t \to v - \phi'(y)z$, 因此在 $H_0^1(\Omega)$ 中 $z_t \to z$, z 是 (6.12) 的解. 证得 $u \to y$ 是 Gâteaux 可微的, 沿方向 v 的方向导数是 (6.12) 的解.

(c) 验证 y_u 的 Gâteaux 导数是 u 的连续函数. 如 2.2 节观察到的, 结论可由连续 Gâteaux 可微映射的 Fréchet 可微性得到.

令 $u_i \in H^{-1}(\Omega)$, 相联系的状态是 $y_i, i = 1, 2$. 置 $\psi_i := \phi'(y_i)$, 给定 $v \in H^{-1}(\Omega)$, 令 z_i 是下述方程:

$$-\Delta z_i + \psi_i z_i = v, \quad i = 1, 2$$

的解. 则 $w := z_2 - z_1$ 是

$$-\Delta w + \psi_2 w = (\psi_1 - \psi_2)z_1 \tag{6.14}$$

的解. 由 Sobolev 嵌入, $H_0^1(\Omega) \subset L_q(\Omega)$, $\dfrac{1}{q} = \dfrac{1}{2} - \dfrac{1}{n}$. 因为 $\dfrac{2}{n} + \dfrac{2}{q} = 1$, 由 Hölder 不等式, 方程 (6.14) 可推出

$$\begin{aligned}\int_\Omega |\nabla w(\omega)|^2 \mathrm{d}\omega &\leqslant \int_\Omega |\psi_2(\omega) - \psi_1(\omega)| \cdot |z_1(\omega)| \cdot |w(\omega)| \mathrm{d}\omega \\ &\leqslant \|\psi_2 - \psi_1\|_{\frac{n}{2}} \|z_1\|_q \|w\|_q.\end{aligned}$$

在 $L_2(\Omega)$ 中, 当 $u_2 \to u_1$ 时, 由步 (a) 有 $\|z_1\|_q = \mathrm{O}(\|v\|_{-1,2})$, 由 Lebesgue 定理, $\|\psi_2 - \psi_1\|_{\frac{n}{2}} \to 0$. 于是得

$$\|w\|_{1,2,0}^2 = o(\|w\|_q \|v\|_{-1,2}) = o(\|w\|_{1,2,0} \|v\|_{-1,2}).$$

因此 $\|z_2 - z_1\|_{1,2,0} = o(\|v\|_{-1,2})$. 证得结论. □

6.2.3 强解

二阶椭圆方程的强解定义为二阶导数是可积函数的分布意义下的解. 当讨论二阶最优性条件时, 考虑状态方程的强解是有用途的. 其原因是, 如果 ϕ 是二次连续可微的, $u \in L_s(\Omega), s > \dfrac{1}{2}n$, 则 y_u 是在 Ω 上为连续的状态方程的强解, 且映射 $u \to y_u$ 是从 $L_s(\Omega)$ 到 $C_0(\Omega)$ 的二次连续可微的.

不加证明地给出处理 Poisson 方程的下述结果[2], 然后详细讨论状态方程的强解. 假设 Ω 具有 C^2 光滑的边界.

引理 6.14 *设 $s \in [2, \infty), f \in L_s(\Omega)$, 则 Dirichlet 问题 (6.1) 具有唯一解 $y_u \in W^{2,s}(\Omega) \cap W_0^{1,s}(\Omega)$, 且存在 $c_s > 0$, 不依赖于 f, 满足*

$$\|y\|_{2,s} \leqslant c_s \|f\|_s. \tag{6.15}$$

用上述引理, 可以相对容易地研究半线性的椭圆方程.

命题 6.15 *设 ϕ 是 Lipschitz 连续且连续可微的, 若 $u \in L_s(\Omega), s \in [2, \infty)$, 则下述结论成立:*

(i) 半线性方程 (6.11) 具有唯一解 $y_u \in W^{2,s}(\Omega) \cap W_0^{1,s}(\Omega)$.

(ii) 若 $s \geqslant \dfrac{1}{2}$, 则

$$\|y\|_{2,s} \leqslant 2c_s \|u\|_s, \tag{6.16}$$

其中 c_s 是引理 6.14 中的常数.

(iii) 若 ϕ 是二次连续可微的, 且 $s > \dfrac{1}{2}n$(若 $n \leqslant 3$ 则 $s = 2$), 则映射 $u \mapsto y_u$ 是从 $L_s(\Omega)$ 到 $W^{2,s}(\Omega) \cap W_0^{1,s}(\Omega)$ 的二次连续可微的映射.

证明 (i) 由命题 6.12, (6.11) 具有唯一解 $y_u \in H_0^1(\Omega)$. 用下述的自助 (bootstrapping) 方法证明 $y_u \in W^{2,s}(\Omega)$, 这一证明基于 Sobolev 包含. 考虑序列 $\{s_i\}$, 其

中 $s_1 := 2$,

$$\begin{cases} 1/s_{i+1} = 1/s_i - 2/n, & 若\ 1/s_i - 2/n > 0, \\ s_{i+1} = +\infty, & 否则. \end{cases} \tag{6.17}$$

当 i 充分大时这一非减序列等于 $+\infty$. 由 Sobolev 嵌入, 除了 $1/s_i - 2/n = 0$ 的情况, 有 $W^{2,s_i}(\Omega) \subset L_{s_{i+1}}(\Omega)$. 对所有的 $\hat{s} < +\infty, W^{2,s_i}(\Omega) \subset L_{\hat{s}}(\Omega)$.

置 $y = y_u$. 因为 ϕ 是 Lipschitz 连续的, 对 $i = 1$ 时, 有 $\Delta y = \phi(y) - u \in L_{s_i}(\Omega)$. 假设对给定的 $i \geqslant 1$ 是成立的, 则由引理 6.14 与 Sobolev 嵌入, 若 $s_{i+1} \leqslant s$, 则 $y \in L_{s_{i+1}}(\Omega)$, 否则 $y \in L_s(\Omega)$. 由归纳法, $\Delta y_u = \phi(y_u) - f \in L_s(\Omega)$, 从而根据引理 6.14, $y_u \in W^{2,s}(\Omega) \cap W_0^{1,s}(\Omega)$.

(ii) 由于 $y_u \in W^{2,s}(\Omega)$ 且 $s > \frac{1}{2}n$, 有 $y_u \in C_0(\Omega)$. 令 $F : \mathbb{R} \to \mathbb{R}$ 定义为 $F(t) := |\phi(t)|^{s-2}\phi(t)$. 由于 F 在有界集上是 Lipschitz 连续的且连续可微的, $y_u \in C_0(\Omega)$, 作为引理 6.11 的推论, 有 $F(y) \in W_0^{1,s}(\Omega)$ 且

$$\nabla F(y) = (s-1)|\phi(y)|^{s-2}\phi'(y)\nabla y.$$

将状态方程乘以 $F(y)$ 并在 Ω 上取积分 (注意积分是有定义的), 用 Green 公式与 Hölder 不等式, 得到

$$\begin{aligned} \Delta := (s-1)\int_\Omega |\phi(y(\omega))|^{s-2}\phi'(y(\omega))|\nabla y(\omega)|^2 \mathrm{d}\omega + \int_\Omega |\phi(y(\omega))|^s \mathrm{d}\omega \\ \leqslant \int_\Omega |u(\omega)| \cdot |\phi(y(\omega))|^{s-1}\mathrm{d}\omega \leqslant \|u\|_s \|\phi(y)^{s-1}\|_{s'} = \|u\|_s \|\phi(y)\|_s^{s-1}. \end{aligned}$$

于是得到 $\|\phi(y)\|_s \leqslant \|u\|_s$, 从而推出 $\|\Delta y\|_s \leqslant 2\|u\|_s$. 包含关系由引理 6.14 得到.

(iii) 步 (a). 由于 $\phi : \mathbb{R} \to \mathbb{R}$ 是二次连续可微的, 映射 $y \mapsto \phi(y) : C_0(\Omega) \to C_0(\Omega)$ 满足

$$\phi(y+z) = \phi(y) + \phi'(y)z + \frac{1}{2}\phi''(y)z^2 + r(z),$$

其中

$$r(z) := \int_0^1 (1-\sigma)[\phi''(y+\sigma z) - \phi''(y)]z^2 \mathrm{d}\sigma.$$

由于 ϕ'' 在 $\mathbb{R}$ 的有界子集上是一致连续的, 可得 (对给定的 y) $r(z) = o(\|z\|_\infty^2)$, 因此 $y \mapsto \phi(y)$ 在 $C_0(\Omega)$ 内是二次连续可微的.

步 (b). 因为 $s > \frac{1}{2}n$, 由 (6.5) 有 $W_0^{2,s}(\Omega) \subset C_0(\Omega)$, 则映射

$$F(y,u) := -\Delta y + \phi(y) - u$$

是 $W^{2,s} \cap W_0^{1,s}(\Omega) \times L_s(\Omega) \to L_s(\Omega)$ 的二次连续可微映射. 还有 $D_y F(y,u)z = -\Delta z + \phi'(y)z$. 由命题 6.13 与自助 (bootstrapping) 方法的推证, 对给定的 $v \in L_s(\Omega)$,

方程 (6.12) 有唯一解 $z \in W^{2,s} \cap W_0^{1,s}(\Omega)$. 由隐函数定理5.14 可以导出映射 $u \mapsto y_u$ 是从 $L_s(\Omega)$ 到 $W^{2,s}(\Omega)$ 的 C^2 光滑映射. □

已经证明了映射 $u \mapsto y_u$ 可微性的两个不同的结论. 设 ϕ 是 Lipschitz 连续的且连续可微的, 我们总是做这样的假设, $u \mapsto y_u$ 是从 $H^{-1}(\Omega)$ 到 $H_0^1(\Omega)$ 的连续可微映射. 若还有 ϕ 是二次可微的, 且 $s > \frac{1}{2}n$, 则 $u \mapsto y_u$ 是从 $L_s(\Omega)$ 到 $W^{2,s}(\Omega) \cap W_0^{1,s}(\Omega)$ 的二次连续可微映射.

6.3 半线性的椭圆方程的最优控制

6.3.1 解的存在性, 一阶最优性系统

考虑二次目标函数:

$$J_\tau(y,u) := \frac{1}{2}\int_\Omega (y(\omega) - \tau(\omega))^2 \mathrm{d}\omega + \frac{1}{2}N\int_\Omega u(\omega)^2 \mathrm{d}\omega, \tag{6.18}$$

其中 $N > 0$ 是给定的, 靶函数 τ 属于 $L_q(\Omega), q \geqslant 2$. 可行控制集合 K 是给定的非空的闭的$L_2(\Omega)$ 中的凸子集. 这样集合的例子将在后面讨论, 见例 6.24. 我们讨论最优控制问题:

$$(\mathrm{P}_\tau) \qquad \min_u \{F(u) := J_\tau(y(u), u)\} \quad \text{s. t.} \quad u \in K. \tag{6.19}$$

令 $\mathcal{S}(\mathrm{P}_\tau)$ 记 (P_τ) 的 (最优) 解集合, $\mathrm{val}(\mathrm{P}_\tau)$ 记 (P_τ) 的最优值, 即 $\mathrm{val}(\mathrm{P}_\tau) := \inf_{u\in K} J_\tau(y(u), u)$.

在这一问题中, 控制空间是 $L_2(\Omega)$. 所以, (P) 的解或局部解指的是 $L_2(\Omega)$ 中的解或局部解. 后面讨论 $s > 2$ 时在 $L_s(\Omega)$ 中的局部解.

定理 6.16 *问题(P_τ)(至少) 有一解.*

证明 由于 K 是非空的, 而 $N > 0$, 问题 (P_τ) 的极小化序列$\{u_k\}$ 是存在的, 且在 $L_2(\Omega)$ 内有界. 根据命题 6.15, 相联系的状态 $\{y_k\}$ 在 $H_0^1(\Omega)$ 中是有界的. 如有必要可抽取一子列, 不妨设 $u_k \xrightarrow{w} \bar{u}$ 于 $L_2(\Omega)$ 中, $y_k \xrightarrow{w} \bar{y}$ 于 $H_0^1(\Omega)$ 中, 因此 $y_k \to \bar{y}$ 于 $L_2(\Omega)$ 中, 从而 $y_k \to \bar{y}$ 几乎处处于 Ω 上.

由引理 6.11, $\phi(y_k)$ 在 $H_0^1(\Omega)$ 中是有界的. 如有必要, 可抽取一子列, 不妨设 $\phi(y_k)$ 在 $H_0^1(\Omega)$ 中是弱收敛的, 因而它在 $L_2(\Omega)$ 中是强收敛的. 由于 $\phi(y_k) \to \phi(\bar{y})$ 几乎处处成立, 有 $\phi(y_k) \xrightarrow{w} \phi(\bar{y})$ 于 $H_0^1(\Omega)$ 中.

对关系 $-\Delta y_k + \phi(y_k) = u_k$ 在 $H^{-1}(\Omega)$ 取弱极限, 得到 $\bar{y}$ 是与 $\bar{u}$ 相联系的状态. 由于 J_τ 是 $L_2(\Omega) \times L_2(\Omega) \to \mathbb{R}$ 的凸的连续函数, 因而是弱下半连续的, 得到

$$J_\tau(\bar{y}, \bar{u}) \leqslant \lim_{k\to\infty} J_\tau(y_k, u_k) = \mathrm{val}(\mathrm{P}_\tau).$$

最后, 由于 K 是 $L_2(\Omega)$ 中的凸闭子集, 因而是弱闭的, 有 $\bar{u} \in K$. 所以 $\bar{u}$ 是 (P_τ) 的一个解. □

注 6.17　上述证明表明, 映射 $u \mapsto y_u$ 是从 $H^{-1}(\Omega)$ 到 $H_0^1(\Omega)$ 的连续映射, 其中这两个空间都赋予弱拓扑.

伴随方程定义为

$$\begin{cases} -\Delta p + \phi'(y_u)p = y_u - \tau, & \text{在 } \Omega \text{ 内}, \\ p = 0, & \text{在 } \partial\Omega \text{ 上}. \end{cases} \tag{6.20}$$

引理 6.18　令 $u \in L_2(\Omega)$, 则下述命题成立:

(i) 伴随方程有唯一的解 $p_u \in H_0^1(\Omega)$, 称为与 u 相联系的伴随状态. 函数 $u \mapsto F(u)$ 是从 $L_2(\Omega)$ 到 $\mathbb{R}$ 的 C^1 光滑函数, 其导数为

$$DF(u) = p_u + Nu. \tag{6.21}$$

(ii) 若 ϕ 是 C^2 类的, u 属于 $L_s(\Omega)$, 其中 $s > \frac{1}{2}n$(若 $n \leqslant 3$, 则 $s = 2$), 则 $F(u)$ 是从 $L_s(\Omega)$ 到 $\mathbb{R}$ 的 C^2 光滑函数, 其导数由 (6.21) 给出.

证明　(i) 由已经给出的工具可以容易证得 (6.20) 在 $H^1(\Omega)$ 中具有唯一解. 由命题 6.13, 作为连续可微映射的复合, 函数 $F(u) := J_\tau(u, y_u) : L_2(\Omega) \to \mathbb{R}$ 是连续可微的, 沿方向 $v \in L_2(\Omega)$ 的导数为

$$DF(u)v = \int_\Omega (y_u - \tau)(\omega)z(\omega)\mathrm{d}\omega + N\int_\Omega u(\omega)v(\omega)\mathrm{d}\omega,$$

其中 z 是 (6.12) 的解. 另一方面, 由 Green 公式, 得

$$\begin{aligned} \int_\Omega (y_u - \tau)(\omega)z(\omega)\mathrm{d}\omega &= \int_\Omega (-\Delta p_u + \phi'(y_u)p_u)z(\omega)\mathrm{d}\omega \\ &= \int_\Omega (-\Delta z + \phi'(y_u)z)p_u(\omega)\mathrm{d}\omega \\ &= \int_\Omega p_u(\omega)v(\omega)\mathrm{d}\omega. \end{aligned}$$

于是得到 $DF(u)v = \int_\Omega (p_u + N_u)(\omega)v(\omega)\mathrm{d}\omega$, 即 $DF(u) = p_u + Nu$.

(ii) 由 (i)、命题 6.15(iii) 与 $J_\tau(\cdot,\cdot)$ 是从 $L_2(\Omega) \times L_2(\Omega)$ 到 $\mathbb{R}$ 的 C^2 类映射即得结论. □

下面叙述 (P_τ) 的一阶最优性条件与它的一个有用的刻画.

命题 6.19　设 u 是 (P_τ) 的一个局部解, 则它满足一阶最优性条件:

$$0 \in DF(u) + N_K(u), \tag{6.22}$$

或等价地,

$$u = P_K(-N^{-1}p_u), \tag{6.23}$$

其中 P_K 是 $L_2(\Omega)$ 中到 K 上的直交投影.

证明 包含关系 (6.22) 是约束函数 G 为单位映射时的标准一阶最优性系统. 设 (6.22) 成立, 令 $v \in K$, 则 $\langle u + N^{-1}p_u, v - u\rangle_2 \geqslant 0$, 所以

$$\|v + N^{-1}p_u\|_2^2 = \|u + N^{-1}p_u\|_2^2 + 2\langle u + N^{-1}p_u, v - u\rangle_2 + \|v - u\|_2^2 \geqslant \|u + N^{-1}p_u\|_2^2.$$

可推出 (6.23). 相反的推出关系可由上述关系很容易得到. □

尽管由函数 J_τ 的形式所示, 选取 $L_2(\Omega)$ 作为控制空间有些任意性. 例如, 如果 $K \subset L_\infty(\Omega)$, 则可以把控制空间取为任何 $L_s(\Omega)$ 空间, $s \in [2, \infty]$. 当 ϕ 是 C^2 类的且 $\frac{1}{2}n < s < \infty$(若 $n \leqslant 3, s = 2$), 这具有确定的优越性, 因为在此情况下, 目标函数 $F(u)$ 是二次可微的, 即可采用二阶分析学.

讨论类似于 (P_τ) 的但用控制空间 $L_s(\Omega)$ 的问题, 其中 $s > 2$, 即

$$(\mathrm{P}_{\tau,s}) \qquad \min_u \{F(u) := J_\tau(u, y_u)\} \quad \text{s. t.} \quad u \in K, \quad u \in L_s(\Omega). \tag{6.24}$$

像以前那样, 这里 K 是 $L_2(\Omega)$ 中的非空闭凸子集. 我们对讨论 (P_τ) 与 $(\mathrm{P}_{\tau,s})$ 的 (可能是局部的) 解之间的关系尤其感兴趣.

例如, 若 $K \subset L_s(\Omega)$, 则 (P_τ) 与 $(\mathrm{P}_{\tau,s})$ 有相同的解. 然而, 即使这种情况, 这些问题的局部解也可能不是重合的, 因为 $L_s(\Omega) \subset L_2(\Omega)$, 只有前者的局部解是后者的局部解. 下面条件成立的时候, 相反的结论亦是成立的.

引理 6.20 令 $s \geqslant 2$, 设 $K \subset L_\infty(\Omega)$, 或更一般地, 存在 $f \in L_s(\Omega)$ 满足对所有的 $v \in K$, 下述关系成立:

$$|v(\omega)| \leqslant f(\omega), \quad \text{a.e.} \quad \omega \in \Omega. \tag{6.25}$$

则 $(\mathrm{P}_{\tau,s})$ 的任何局部解是 (P_τ) 的局部解.

证明 假设结论不成立, 即存在 $(\mathrm{P}_{\tau,s})$ 的局部解不是 (P_τ) 的局部解. 这意味着存在 $\{u_k\} \subset K$, $u_k \to u$ 于 $L_2(\Omega)$ 中, 满足 $F(u_k) < F(u)$, 则有 $u_k \to u$ a.e., 且根据 (6.25), Lebesgue 定理可推出 $u_k \to u$ 于 $L_s(\Omega)$ 中. 于是 $F(u_k) < F(u)$, 与 u 是 $(\mathrm{P}_{\tau,s})$ 的局部解矛盾. □

现在讨论两个问题最优性条件间的关系. 置

$$K_s := K \cap L_s(\Omega),$$

在空间 $L_s(\Omega)$ 中, 令 $T_{K_s}(u)$ 是 K_s 在 $u \in L_s(\Omega)$ 处的切锥, 即空间 $L_s(\Omega)$ 中的 $\mathcal{R}_K(u) \cap L_s(\Omega)$ 的闭包 (雷达锥$\mathcal{R}_K$ 由 (2.83) 定义). 因为 $L_s(\Omega) \subset L_2(\Omega)$, 给定 $u \in$

K_s, 有 $T_{K_s}(u) \subset T_K(u)$. 由于 $L_s(\Omega)$ 是 $L_2(\Omega)$ 的稠密子集, 极锥满足 $N_K(u) \subset N_{K_s}(u)$. 可得到, (P_τ) 的最优性条件推出 $(\mathrm{P}_{\tau,s})$ 的最优性条件, 即 (6.22) 可推出

$$0 \in DF(u) + N_{K_s}(u). \tag{6.26}$$

下面叙述使两个问题的一阶最优性系统重合的一个条件.

命题 6.21　令 $u \in L_s(\Omega)$ 满足 (6.26), 其中 $s > 0$. 若

$$\mathcal{R}_K(u) \cap L_s(\Omega) \text{ 是 } T_K(u) \text{ 的稠密子集}, \tag{6.27}$$

则 (6.22) 与 (6.23) 成立.

证明　令 $v \in \mathcal{R}_K(u) \cap L_s(\Omega)$, 则 $DF(u)v \geqslant 0$. 因为 $DF(u) \in L_2(\Omega)$, 由 (6.27) 可以推出, 对所有的 $v \in T_K(u)$, 均有 $DF(u)v \geqslant 0$, 即 (6.22) 成立. 与命题 6.19 相结合, 得到 (6.23). □

注 6.22　若 $K \subset L_s(\Omega)$, 或更一般地, 若 $K \cap L_s(\Omega)$ 是 K 的稠密子集, 则 (6.27) 成立.

下面给出的 (某种程度病态的) 例子表明, 条件 (6.22) 与 (6.26) 不总是等价的.

例 6.23　令 $s > 2, u_1 \in L_s(\Omega), u_2 \in L_2(\Omega) \setminus L_s(\Omega), K = [u_1, u_2]$ (u_1 与 u_2 的凸组合的集合). 则 $u = u_1$ 是 $(\mathrm{P}_{\tau,s})$ 的唯一解. 由于 $\mathcal{R}_K(u_1) \cap L_s(\Omega) = \{0\}$, 很清楚, (P_τ) 与 $(\mathrm{P}_{\tau,s})$ 的一阶最优性条件是不等价的.

例 6.24　考虑常界值约束, 即 K 具有下述形式的情况:

$$K_{a,b} := \{u \in L_2(\Omega): \quad a \leqslant u(\omega) \leqslant b \ \text{ a.e. } \ \omega \in \Omega\}, \tag{6.28}$$

其中 $-\infty \leqslant a < b \leqslant +\infty$. 容易验证, 此种情况下 (6.27) 是成立的. 事实上, 由于 $\mathcal{R}_K(u)$ 是 $T_K(u)$ 的稠密子集, 只需要观察到, 当 $M \uparrow +\infty$ 时, 方向 $v \in \mathcal{R}_K(u)$ 是截断函数

$$[v]_M(\omega) := \max\{-M, \min\{M, v(\omega)\}\}, \quad \text{a.e. } \ \omega \in \Omega \tag{6.29}$$

在 $L_2(\Omega)$ 中的极限; 同时, $[v]_M$ 属于 $\mathcal{R}_K(u) \cap L_\infty(\Omega)$.

在 $L_2(\Omega)$ 中到 $K_{a,b}$ 上的投影由下述公式给出:

$$P_{K_{a,b}}(u)(\omega) = \max(a, \min(u(\omega), b)), \quad \text{a.e. } \ \omega \in \Omega. \tag{6.30}$$

注意到, 若 $u \in L_s(\Omega)$, 则在空间 $L_s(\Omega)$ 中, u 到 $K_{a,b} \cap K_s(\Omega)$ 上的投影是有定义的, 且与它在 $L_2(\Omega)$ 中到 $K_{a,b}$ 上的投影是重合的.

因为 $p_u \in H^2(\Omega) \cap H_0^1(\Omega)$, 若 $n \leqslant 3$, 则由 Sobolev 嵌入, p_u 属于 $C_0(\Omega)$, 因此由 (6.23), u 也属于 $C_0(\Omega)$. 若 $n > 3$, 则 p_u 属于 $L_{s_2}(\Omega)$, 其中 $1/s_2 := 1/2 - 2/n$. 根据 (6.23) 与 (6.30), 可推出下述正则性结果: (P_τ) 的局部解属于 $L_{s_2}(\Omega)$. 用自助法

推证可以得到更多的结论. 例如, 下述结果给出了推出状态的连续性的靶函数 τ 上的条件.

引理 6.25 设 u 是 (P_τ) 的局部解, 在 $L_2(\Omega)$ 中到 K 上的直交投影将 $L_s(\Omega)$ 映到 $L_s(\Omega), \forall s \in [2, \infty)$(若 K 是由 (6.28) 定义的 $K_{a,b}$ 的形式, 就是这种情况). 令 $\tau \in L_q(\Omega)$, 其中 $q > \max\left(2, \frac{1}{4}n\right)$ (若 $n < 8$, 则 $q = 2$), 则 y_u 是连续函数.

证明 首先观察到, 因为 $\phi'(y_u) \in L_\infty(\Omega)$, 由自助法推证, 若 $y_u \in L_r(\Omega)$, 其中 $2 \leqslant r \leqslant q$, 则 $p_u \in W^{2,r}(\Omega)$.

用到 (6.17) 定义的序列 $\{s_i\}$, 再次用自助法推证. 当 $u \in L_{s_i}(\Omega), i = 1$ 时就是这样的情况, 有 $y_u \in W^{2,s_i}(\Omega) \subset L_{s_{i+1}}(\Omega)$. 只要 $s_{i+1} \leqslant q$, 就有 $p_u \in W^{2,s_{i+1}}(\Omega) \subset L_{s_{i+2}}(\Omega)$, 因此 $u \in L_{s_{i+2}}(\Omega)$. 由归纳法可得 $p \in W^{2,q}(\Omega)$, 因此 $u \in L_s(\Omega), 1/s = 1/q - 2/n$. 由于 $q > \frac{1}{4}n$, 有 $s > \frac{1}{2}n$, 则由命题 6.15 与 Sobolev 嵌入, 可得到状态的连续性. □

注 6.26 用命题 6.45 建立的 $W^{1,s}(\Omega)$ 的 Banach 格结构, 在 $K = K_{a,b}$ 的情况下, 辅加的正则性结果是可以得到的, 即若 $p_u \in W_0^{1,s}(\Omega)$, 则 $u \in W_0^{1,s}(\Omega)$.

6.3.2 二阶必要或充分性条件

这一节, 设 ϕ 是二阶连续可微的. 首先给出此种情况下目标函数 Hesse 阵的一个表达式. Banach 空间X 上的二次型 Q 是 Legendre 形式(见定义 3.73): 若 Q 是弱下半连续的, 序列 $\{x_k\} \subset X$ 满足 $x_k \xrightarrow{w} x$ 且 $Q(x_k) \to Q(x)$, 则 $x_k \to x$.

引理 6.27 设 $s > \frac{1}{2}n$ 且 ϕ 是二阶连续可微的, 则下述结论成立:

(i) 函数 $F(u) := J_\tau(y_u, u) : L_s(\Omega) \to \mathbb{R}$ 是二次连续可微的, $F(\cdot)$ 在 $u \in L_s(\Omega)$ 处沿方向 $v \in L_s(\Omega)$ 的二阶导数是

$$D^2F(u)(v,v) = \int_\Omega [Nv(\omega)^2 \mathrm{d}\omega + (1 - p_u(\omega)\phi''(y_u(\omega)))z_v(\omega)^2]\mathrm{d}\omega, \tag{6.31}$$

其中 z_v 是 (6.12) 的解.

(ii) 若还有 $\tau \in L_q(\Omega), q > \frac{1}{2}n$, 则 $D^2F(u)$ 在 $L_2(\Omega) \times L_2(\Omega)$ 上有唯一的连续的延拓, 且后者是 Legendre 形式.

证明 (i) 由命题 6.15, $F(u)$ 是二次连续可微映射的复合函数, 因此是二次连续可微的. Hesse 阵的表达式可通过考虑 Lagrange 函数:

$$L(y,u,p) := J_\tau(y,u) - \int_\Omega p(\omega)(-\Delta y + \phi(y) - u)(\omega)\mathrm{d}\omega$$

得到. 因为 $F(v) = L(y_v, v, p_v), D_yL(y_u, u, p_u) = 0$, 有

$$F(u+v) = F(u) + DF(u)v + \frac{1}{2}D^2_{(y,u)^2}L(y_u,u,p_u)((z_v,v),(z_v,v)) + o(\|v\|_s^2),$$

其中 z_v 是线性化状态方程 (6.12) 的解. 结果得证.

(ii) 由 (6.31), 有

$$D^2F(u)(v,v) = N\|v\|_2^2 + Q_u(v), \tag{6.32}$$

其中 $Q_u(v)$ 表示 F 的 Hesse 阵的状态分布. 若 $q > \frac{1}{2}n$, 则 $p \in W^{2,s^*}(\Omega)$, 其中 $s^* := \min(s,q) > \frac{1}{2}n$. 所以, $p_u\phi''(y_u) \in C_0(\Omega)$. 因为由命题 6.13, $v \mapsto z_v : L_2(\Omega) \to L_2(\Omega)$ 是紧致的, 由 (i), $Q_u(v)$ 在 $L_2(\Omega)$ 上有唯一的到弱连续的二次形式的延拓. 显然, $N\|v\|_2^2$ 也有唯一的 Legendre 形式的延拓. 作为 Legendre 形式与弱的连续的二次形式的和, $D^2F(u)$ 自身是 Legendre 形式. □

上述结果启发我们, 对于 $s > \frac{1}{2}n$, 取 $L_s(\Omega)$ 作为控制空间 (若 $n \leqslant 3$, 则 $s = 2$). 若 $n > 3$, 对于 $s > \frac{1}{2}n$, (P_τ) 在 $L_s(\Omega)$ 中解的存在性不能被定理 6.16 保证. 然而, 若 K 是 $L_s(\Omega)$ 的有界子集, 通过稍微修正证明, 可以验证, (P_τ) 在 $L_s(\Omega)$ 中有解. 也可能发生对于某一 $s > \frac{1}{2}n$, $L_2(\Omega)$ 中的解属于 $L_s(\Omega)$ 的情形. 当 $K = K_{a,b}$ 且 $\tau \in L_q(\Omega), q > \frac{1}{4}n$ 时, 由引理 6.25 的证明可得到这一结果.

下面讨论二阶最优性条件. 在 $L_2(\Omega)$ 与 $L_s(\Omega)$ 中临界锥的表达式分别为

$$\begin{aligned} C(u) &:= \{v \in L_2(\Omega) : DF(u)v = 0; v \in T_K(u)\}, \\ C_s(u) &:= \{v \in L_s(\Omega) : DF(u)v = 0; v \in T_{K_s}(u)\}. \end{aligned} \tag{6.33}$$

s 多面性意味着 K_s 在 $L_s(\Omega)$ 中的多面性. 换言之, K 在 $u \in K_s$ 处是 s 多面性的, 若对所有的 $u^* \in N_{K_s}(u)$, 集合 $\mathcal{R}_K(u) \cap (u^*)^\perp$($L_s(\Omega)$ 的子集) 在 $C_s(u)$ 中是稠密的. 还用到下述假设

$$C_s(u) \text{ 是 } C(u) \text{ 的稠密子集.} \tag{6.34}$$

在 6.3.3 节中将讨论这些条件成立的几个例子.

在下面叙述二阶必要条件的时候区分几种情况：先是问题$(\mathrm{P}_{\tau,s})$ 的“标准”条件, 然后是 $L_2(\Omega)$ 中所有临界方向成立的条件. 对于后者, 需要 τ 在 $L_q(\Omega)$ 中, 其中 $q > \frac{1}{2}n$.

定理 6.28　令 u 是 $(\mathrm{P}_{\tau,s})$ 的局部解, 其中 $s > \frac{1}{2}n$, 则下述结论成立:

(i) 若 K 是 s 多面性的, 则对所有的 $v \in C_s(u)$, 下述不等式成立:

$$D^2F(u)(v,v) \geqslant 0. \tag{6.35}$$

(ii) 设 $\tau \in L_q(\Omega)$, 其中 $q > \frac{1}{2}n$ (若 $n \leqslant 3$ 时 $q = 2$), K 是 s 多面性的, 且 (6.34) 是成立的, 则对所有的 $v \in C(u)$, (6.35) 成立.

证明 (i) 令 $v \in \mathcal{R}_K(u) \cap DF(u)^{\perp} \cap L_s(\Omega)$. 则对 $t > 0$ 充分小, 由 $u + tv$ 是可行的, 且 $DF(u)v = 0$, 有

$$0 \leqslant 2 \lim_{t \downarrow 0} t^{-2}(F(u+tv) - F(u)) = D^2F(u)(v,v).$$

由 s 多面性, 这样的 v 的集合在 $C_s(u)$ 中是稠密的, 而 $v \mapsto D^2F(u)(v,v)$ 是连续函数 $L_s(\Omega) \to \mathbb{R}$, 二阶必要性条件成立.

(ii) 由于 $q > \frac{1}{2}n$(若 $n \leqslant 3$ 时 $q = 2$), 由引理 6.27(ii), 映射 $v \mapsto D^2F(u)(v,v)$ 有从 $L_s(\Omega)$ 到 $L_s(\Omega)$ 的连续延拓. 结果由结论 (i) 与 (6.34) 可得到. □

现在转到最优性的充分性条件的研究. 以 $L_2(\Omega)$ 范数的二阶增长条件的表达式成立, 若存在 $\alpha > 0$, 下述条件成立

$$F(v) \geqslant F(u) + \alpha\|v-u\|_2^2 + o(\|v-u\|_2^2), \quad \forall v \in K. \tag{6.36}$$

基于 $L_2(\Omega)$ 范数, 下述表达式出现在二阶充分性最优性条件的讨论中:

$$\exists \alpha > 0, \quad \forall v \in C(u), \quad D^2F(u)(v,v) \geqslant \alpha\|v\|_2^2. \tag{6.37}$$

$$\exists v \in C(u)\backslash\{0\}, \quad D^2F(u)(v,v) > 0. \tag{6.38}$$

先讨论 $n \leqslant 3$ 的情况.

定理 6.29 设 $n \leqslant 3$. 若 K 是多面性的, 则二阶增长条件(6.36) 与二阶充分性条件 (6.37) 和 (6.38) 是等价的.

证明 由于 K 是多面性的, D^2F 是 Legendre 形式, 这是命题 3.53、定理 3.63 与命题 3.74 的一个简单结果, 这里就不给出细节了, 因为一个直接的证明可从定理6.31 的证明中抽取. □

将给出当 $n > 3$ 时上述定理的一推广, 它基于下述的结果:

引理 6.30 令 $s > \frac{1}{2}n$ 与 $q > \frac{1}{2}n$ (若 $n \leqslant 3$ 则 $s = q = 2$). 在 $L_s(\Omega)$ 中给定 u 与 v, 令 $r(u,v)$ 为 F 的二阶展开式中的余项, 即

$$F(u+v) = F(u) + DF(u)v + \frac{1}{2}D^2F(u)(v,v) + r(u,v). \tag{6.39}$$

设 $\tau \in L_q(\Omega)$, 则若 $v \to 0$ 于 $L_2(\Omega)$ 中且 v 在 $L_s(\Omega)$ 中的有界子集中, 有 $|r(u,v)|/\|v\|_2^2 \to 0$.

证明 由命题 6.15, $u \mapsto F(u)$ 是 $L_s(\Omega) \to \mathbb{R}$ 的 C^2 类的, 则 (6.39) 成立, 其中

$$r(u,v) = \int_0^1 (1-\sigma)[D^2F(u+\sigma v) - D^2F(u)](v,v)\mathrm{d}\sigma.$$

记 z_v 为 (6.12) 的解, $z_{\sigma,v}$ 为 $y=y_{u+\sigma v}$ 时相应的方程的解. 根据 (6.31), $r(u,v)$ 等于

$$\int_0^1(1-\sigma)\int_\Omega[(1-p_{u+\sigma v}(\omega)\phi''(y_{u+\sigma v}(\omega)))z_{\sigma,v}^2(\omega)-(1-p_u(\omega)\phi''(y_u(\omega)))z_v^2(\omega)]\mathrm{d}\omega\mathrm{d}\sigma.$$

因此

$$\begin{aligned}|r(u,v)|\leqslant&\frac{1}{2}\sup_{\sigma\in[0,1]}\int_\Omega|(1-p_u(\omega)\phi''(y_u(\omega)))(z_v^2(\omega)-z_{\sigma,v}^2(\omega))|\mathrm{d}\omega\\&+\frac{1}{2}\sup_{\sigma\in[0,1]}\int_\Omega|(p_u(\omega)\phi''(y_u(\omega))-p_{u+\sigma v}(\omega)\phi''(y_{u+\sigma v}(\omega))z_{\sigma,v}^2(\omega))|\mathrm{d}\omega.\end{aligned}$$

由 $\|v-u\|_s$ 是有界的, $\|v-u\|_2\to 0$, 有 $y_{u+\sigma v}$ 与 $p_{u+\sigma v}$ 在 $L_\infty(\Omega)$ 范数意义下是有界的, 按 $H^1(\Omega)$ 的范数一致收敛到 y_u 与 p_u(对 $\sigma\in(0,1)$), 因此是几乎处处收敛的. 于是, 对任何 $t\in(1,+\infty)$, 在 $L_t(\Omega)$ 中, $p_{u+\sigma v}\phi''(y_{u+\sigma v})\to p_u\phi''(y_u)$. 由于在 $H^1(\Omega)$ 中 $z_{\sigma,v}^2\to z_v^2$, 关于 $\sigma\in(0,1)$ 是一致的, 因此, 对 $t>2$, 由 $L_t(\Omega)$ 中的 Sobolev 包含, 用 Hölder 不等式可得上述积分的第二项是 $o(\|z_{\sigma,v}\|_{1,2}^2)$ 阶的. 所以,

$$|r(u,v)|\leqslant O(\|z_v^2-z_{\sigma,v}^2\|_1)+o(\|z_{\sigma,v}\|_2^2).\tag{6.40}$$

因为 $z_v-z_{\sigma,v}$ 是方程:

$$-\Delta(z_v-z_{\sigma,v})+\phi'(y_u)(z_v-z_{\sigma,v})=(\phi'(y_{u+\sigma v})-\phi'(y_u))z_{\sigma,v},\quad 在\ \Omega\ 内,$$

的解. 它在 $\partial\Omega$ 上取 0 值, 再次用到对所有的 $t\in(2,\infty)$ 有 $\|\phi'(y_{u+\sigma v})-\phi'(y_u)\|_t\to 0$ 与 Hölder 不等式, 则上式右端的范数是 $o(\|z_{\sigma,v}\|_{1,2})$ 阶的. 所以,

$$\|z_v-z_{\sigma,v}\|_2=o(\|z_{\sigma,v}\|_{1,2})=o(\|v\|_2).\tag{6.41}$$

由 Cauchy-Schwartz 不等式, 可得到

$$\begin{aligned}\|z_v^2-z_{\sigma,v}^2\|_1&=\int_\Omega|z_v(\omega)+z_{\sigma,v}(\omega)|\cdot|z_v(\omega)-z_{\sigma,v}(\omega)|\mathrm{d}\omega\\&\leqslant\|z_v+z_{\sigma,v}\|_2\cdot\|z_v-z_{\sigma,v}\|_2=o(\|v\|_2^2).\end{aligned}$$

与 (6.40) 相结合, 可以得到结果. □

上述结果建议我们考虑下述二阶增长条件, 它类似于在 $L_2(\Omega)$ 中的条件, 但是在 $L_s(\Omega)$ 中的局部化意义的: 存在 $\alpha>0$, 和在 $L_s(\Omega)$ 中的 u 的某一邻域 $\mathcal{V}$, 成立

$$F(v)\geqslant F(u)+\alpha\|v-u\|_2^2+o(\|v-u\|_2^2),\quad\forall v\in K\cap\mathcal{V}.\tag{6.42}$$

若 $n\leqslant 3$, 则下述结论简化到定理 6.29.

定理 6.31 设 $K \subset L_s(\Omega)$, $\tau \in L_q(\Omega)$, $s > \frac{1}{2}n$ 且 $q > \frac{1}{2}n$(若 $n \leqslant 3$, $s = q = 2$). 令 $u \in F(P_\tau)$. 若 K 是 s 多面性的, (6.34) 成立, 则二阶增长条件 (6.42)、二阶充分性条件 (6.37) 与关系 (6.38) 均是等价的.

证明 证明 (6.42)⇒(6.37)⇒(6.38)⇒(6.42). 若 (6.42) 成立, 令 $v \in \mathcal{R}_K(u) \cap (DF(u))^{\perp} \cap L_s(\Omega)$. 根据 (6.42),

$$F(u+tv) - F(u) \geqslant \alpha t^2 \|v\|_2^2 + o(t^2),$$

因此 $D^2F(u)(v,v) \geqslant 2\alpha\|v\|_2^2$. 由 K 的 s 多面性与 (6.34) 可得二阶充分性条件.

由 (6.37) 推出 (6.38) 是平凡的. 现在设 (6.38) 成立, 而 (6.42) 不成立, 则在 $L_s(\Omega)$ 中存在 $u_k \to u$, 满足

$$F(u_k) < F(u) + \frac{1}{k}\|u_k - u\|_2^2. \tag{6.43}$$

如有必要可抽取一子列, 表示 $u_k = u + t_k v_k$, $t_k \in \mathbb{R}_{++}$, $t_k \downarrow 0$, $\|v_k\|_2 = 1$[①], 在 $L_2(\Omega)$ 中 $v_k \xrightarrow{w} \bar{v}$. 用 (6.43) 与二阶展开式 (6.39), 得到 $DF(u)\bar{v} = 0$(因为 $DF(u)v_k \geqslant 0$) 且 $\limsup\limits_k D^2F(u)(v_k, v_k) \leqslant 0$. 因为 v_k 属于弱闭集 $T_K(u)$, 所以 $\bar{v}$ 是临界方向. 由引理 6.27 可得 $D^2F(u)(\cdot,\cdot)$ 是下半连续的, $D^2F(u)(\bar{v},\bar{v}) \leqslant 0$. 与 (6.32) 相结合, 得到

$$Q_u(\bar{v}) = \lim_{k\to\infty} Q_u(v_k) \leqslant -N \liminf_{k\to\infty} \|v_k\|_2^2 = -N. \tag{6.44}$$

且 $\bar{v}$ 不可能是 0. 但不等式 $D^2F(u)(\bar{v},\bar{v}) \leqslant 0$, 与 (6.38) 矛盾. □

注 6.32 为简便起见, 设 $n \leqslant 3$. $v \mapsto D^2F(u)(v,v)$ 是 Legendre 形式. 如果 ϕ 是 C^2 类的, 可将上述的二阶分析同切二次规划问题相联系, 切二次规划问题如下定义:

$$(\mathrm{Q}_\tau) \qquad \min_v DF(u)v + \frac{1}{2}D^2F(u)(v,v); \quad v \in T_K(u). \tag{6.45}$$

设 u 满足一阶最优性系统, 问题 (P_τ) 与 (Q_τ) 具有相同的临界锥且目标函数具有相同的二阶展开式, 切二次规划问题在 $v = 0$ 处总是多面性的. 如果 K 是多面性的, 对于问题(P_τ) 与 (Q_τ) 而言, 它们的最优性的二阶必要 (充分) 条件是重合的. 根据定理6.28, 问题 (P_τ) 的弱二阶增长条件成立当且仅当它对于 (Q_τ) 也成立.

如果 $\bar{u}$ 是 (P_τ) 的局部解, 则也是 (Q_τ) 的局部解, 它是孤立的当且仅当二阶增长条件成立, 即如果二阶增长条件不成立, 则存在切二阶问题(Q_τ) 的非零的解.

---

① 原著中为 $t_k \in \mathbb{R}_{+*}$.

6.3.3 某些具体的控制约束

对某些具体的例子来验证 s 多面性的假设. 先讨论 $K = K_{a,b}$ 的情况, 然后将结果推广到局部线性约束的情况. 置

$$\{u = a\} := \{\omega \in \Omega : \ u(\omega) = a\},$$

对 $\{a < u < b\}$ 等采用相类似的约定. 这些集合在不考虑测度为零的集合的前提下定义.

命题 6.33 对所有的 $s \in [2, \infty)$, 集合 $K_{a,b} \cap L_s(\Omega)$ 是 s 多面性的.

证明 结论由定理 3.58 可得到, 以下给出直接证明. 令 $u \in K_s$, 其中 $K_s := K_{a,b} \cap L_s(\Omega)$, 则

$$T_{K_s}(u) = \{v \in L_s(\Omega) : v(\omega) \geqslant 0 \text{ 于 } \{u = a\} \text{ 上}, \ v(\omega) \leqslant 0 \text{ 于 } \{u = b\} \text{ 上}\}.$$

令 $u^* \in N_{K_s}(u)$, 因而 $u^* \in L_{s'}(\Omega)$, 有 $u^* \leqslant 0$, $u^* = 0$, $u^* \geqslant 0$ 分别于 $\{u = a\}$, $\{a < u < b\}$, $\{u = b\}$ 上成立. 令 $C(u, u^*) := T_K(u) \cap (u^*)^\perp$, 则

$$C(u, u^*) = \{v \in T_{K_s}(u) : v(\omega) u^*(\omega) = 0 \text{ 于 } \Omega \text{ 上几乎处处成立}\}. \tag{6.46}$$

令 $v \in C(u, u^*)$. 根据 Lebesgue 定理, v 是函数

$$v_\varepsilon(\omega) := \varepsilon^{-1}(P_{[a,b]}(u(\omega) + \varepsilon v(\omega)) - u(\omega)) \text{ 在 } \Omega \text{ 上几乎处处成立} \tag{6.47}$$

的极限, 则 $K_{a,b}$ 的 s 多面性成立. □

对每一 ω, 控制变量是 $\mathbb{R}^l$ 的元素, 考虑 K 由下述局部约束定义的情况:

$$K := \{u \in L_s^l(\Omega) : u(\omega) \in \underline{K}(\omega) \text{ 于 } \omega \in \Omega \text{ 几乎处处成立}\}, \tag{6.48}$$

其中 $\underline{K}(\cdot)$ 是于 Ω 上为几乎处处的 $\mathbb{R}^l$ 的闭凸子集.

引理 6.34 令 K 是满足 (6.48) 的 $L_2^l(\Omega)$ 的闭凸子集, 其中 $\underline{K}(\omega)$ 是 $\omega \in \Omega$ 几乎处处的闭凸集合. 令 $u \in L_s^l(\Omega)$ 与 $u^* \in N_{K_s}(u)$, 则对于所有的 $s \in [2, \infty)$, 下述结论成立:

(i) $T_{K_s}(u) = \{v \in L_s^l(\Omega) : \ v(\omega) \in T_{\underline{K}(\omega)}, \ \text{于 } \omega \in \Omega \text{ 几乎处处成立}\}$.

(ii) $N_{K_s}(u) = \{z \in L_{s'}^l(\Omega) : z(\omega) \in N_{\underline{K}(\omega)}, \ \text{于 } \omega \in \Omega \text{ 几乎处处成立}\}$.

(iii) $T_{K_s}(u) \cap (u^*)^\perp = \{v \in T_K(u) : u^*(\omega) \cdot v(\omega) = 0, \ \text{于 } \omega \in \Omega \text{ 几乎处处成立}\}$.

证明 令 $u \in K$, $v \in \mathcal{R}_K(u)$, 则存在 $\varepsilon > 0$, $u + \varepsilon v \in K$, 因为于 Ω 上几乎处处成立 $(u + \varepsilon v)(\omega) \in K(\omega)$, 从而得到 $v(\omega) \in \mathcal{R}_{K(\omega)}(u)$ 几乎处处成立. $L_s(\Omega)$ 的收敛性可推出几乎处处收敛, 任何 $v \in T_K(u)$ 属于 (i) 中的右端 $\widehat{T}_{K_s}(u)$.

相反地, 令 $v \in \widehat{T}_{K_s}(u)$. 置 $v_\varepsilon := \varepsilon^{-1}(P_K(u+\varepsilon v)-u)$, 则在 Ω 上几乎处处有

$$v_\varepsilon(\omega) := \varepsilon^{-1}(P_{K(\omega)}(u(\omega)+\varepsilon v(\omega))-u(\omega))$$

属于 $\mathcal{R}_{K(\omega)}(u)$, $v_\varepsilon(\omega) \to v(\omega)$ 几乎处处成立. 因为 $\mathbb{R}^l$ 内的投影是压缩映射, 所以 $|v_\varepsilon(\omega)| \leqslant |v(\omega)|$ 于 Ω 上几乎处处成立. 根据 Lebesgue 定理可得, 在 $L_s(\Omega)$ 内有 $v_\varepsilon \to v$, 所以有 $\widehat{T}_{K_s}(u) \subset T_{K_s}(u)$, 证得 (i).

令 $\widehat{N}_{K_s}(u)$ 为 (ii) 的右端. 如果 $\lambda \in \widehat{N}_{K_s}(u)$, 则由 (i) 得 $\int_\Omega v(\omega)\cdot\lambda(\omega)\mathrm{d}\omega \leqslant 0$, $\forall v \in T_{K_s}(u)$, 所以有 $\widehat{N}_{K_s}(u) \subset N_{K_s}(u)$. 令 $\lambda \in L^l_{s'}(\Omega)\backslash\widehat{N}_{K_s}(u)$, 则存在 $v \in T_{K_s}(u)$ 与 Ω 的可测子集 Ω_1, 满足 meas$(\Omega_1)>0$ 与 $\alpha>0$, 使得 $\lambda(\omega)\cdot v(\omega)>\alpha$ 于 Ω_1 上几乎处处成立, $|v(\omega)|=1$. 可设 v 在 Ω_1 取常值, 在 $\Omega\backslash\Omega_1$ 上取零值. 由 (i) 得到 $v \in T_{K_s}(u)$, 而 $\int_\Omega v(\omega)\cdot\lambda(\omega)\mathrm{d}\omega>0$, 从而得到 $\lambda \notin N_{K_s}(u)$. 证得 (ii). 最后 (iii) 是 (i) 与 (ii) 的简单结论. □

现在研究 $\underline{K}(\omega)$ 由有限多个线性约束定义的情形.

命题 6.35 令 K 是满足 (6.48) 的 $L^l_2(\Omega)$ 的闭凸子集, 其中

$$\underline{K}(\omega)=\{x\in\mathbb{R}^l:\langle a_i(\omega),x\rangle\leqslant b_i(\omega),\quad i\in J(\omega)\},$$

这里 $J(\omega)$ 是包含在 $\{1,\cdots,q\}$ 的有限集合, $\omega\mapsto J(\omega)$ 是可测映射且 $a_i(\cdot)(b_i(\cdot))$ 属于 $L^l_\infty(\Omega)(L^q_\infty(\Omega))$, $i=1,\cdots,q$, 则对所有的 $s\in[2,\infty)$, $K\cap L_s(\Omega)$ 是 $L^l_s(\Omega)$ 中满足 s 多面性的闭凸子集, 且 (6.34) 成立.

证明 因为 K 是 $L^l_2(\Omega)$ 的闭凸子集, 可以用引理 6.34. 条件 (6.34) 可以用截断性论证证明. 下证 $K\cap L_s(\Omega)$ 是 s 多面性的. 固定 $u\in K$, $u^*\in N_K(u)$, 令 $v\in T_K(u)\cap(u^*)^\perp$. 给定 $\varepsilon>0$, 置

$$v_\varepsilon(\omega)=\begin{cases}v(\omega), & 若\ |v(\omega)|\leqslant\varepsilon^{-1}\ 且\ u(\omega)+\varepsilon v(\omega)\in\underline{K}(\omega),\\ 0, & 否则\end{cases}$$

于 Ω 上是几乎处处的. 则根据引理 6.34, $v\in L_\infty(\Omega)$, 有 $v_\varepsilon\in T_K(u)\cap(u^*)^\perp$, 且 $|v_\varepsilon(\omega)|\leqslant|v(\omega)|$, 由引理 6.34(i), 于 Ω 上几乎处处有 $v_\varepsilon(\omega)\to v(\omega)$. 根据 Lebesgue 定理可得在 $L^l_s(\Omega)$ 内有 $v_\varepsilon\to v$. 结果得证. □

6.3.4 灵敏性分析

假设 $s>\frac{1}{2}n$ 且 $q>\frac{1}{2}n$, 这一节讨论将 (P_τ) 的局部解视为参数 τ 的函数的性态. 先讨论一个稳定性结果. 引入下述记号:

$$F(u,\tau):=J_\tau(y_u,u),\quad \mathrm{val}(\tau):=\inf_{u\in K}F(u,\tau).$$

对 $(u, y_u) = (\bar{u}, \bar{y})$, 用 $\bar{z}_v$ 记 (6.12) 的解, $\bar{p}$ 是与 $(\bar{u}, \bar{y})$ 相联系的伴随状态. 令 $\eta \in L_q(\Omega)$, 则类似于引理 6.27 中所述, 有

$$D^2F(\bar{u}, \bar{\tau})((v, \eta), (v, \eta)) = \int_\Omega (Nv(\omega)^2 + (\bar{z}_v(\omega) - \eta(\omega))^2 - \bar{p}(\omega)\phi''(\bar{y}(\omega))\bar{z}_v(\omega)^2)\mathrm{d}\omega.$$

注意到, 由引理 6.30 可得, $D^2F(\bar{u}, \bar{\tau})$ 从 $L_s(\Omega) \times L_s(\Omega)$ 到 $L_2(\Omega) \times L_2(\Omega)$ 有连续的扩展. 引入如下的子问题:

$$\text{(SP)} \qquad \min_v D^2F(\bar{u}, \bar{\tau})((v, \eta), (v, \eta)) \quad \text{s. t.} \quad v \in C(\bar{u}). \tag{6.49}$$

定理 6.36　设 $s > \frac{1}{2}n$ 与 $q > \frac{1}{2}n$(若 $n \leqslant 3$ 时 $s = q = 2$), K 是 s 多面的, 稠密条件 (6.34) 成立, 且下述假设成立:

$$\text{或者 } n \leqslant 3 \text{ 或者 } K \text{ 是 } L_s(\Omega) \text{ 的有界子集.} \tag{6.50}$$

则下述结论成立:

(i) 若在 $L_s(\Omega)$ 中 $\tau_k \to \bar{\tau}$, $u_k \in S(\mathrm{P}_{\tau_k})$, 则存在 $\bar{u} \in S(\mathrm{P}_{\bar{\tau}})$ 是 $\{u_k\}$ 一子列在 $L_s(\Omega)$ 中的弱极限, 在 $L_2(\Omega)$ 的强极限. 若还有 $\bar{u}$ 满足弱二阶充分性条件 (6.38)(对问题$(\mathrm{P}_{\bar{\tau}})$), 则对相应的子列, 有

$$\|u_k - \bar{u}\|_2 = O(\|\tau_k - \bar{\tau}\|_2). \tag{6.51}$$

(ii) 若 $\tau_k = \bar{\tau} + t_k\eta$, $\eta \in L_q(\Omega)$, 则上述 $\bar{u}$ 是下面的线性化问题的解

$$\text{(LP)} \qquad \min_u \langle \bar{\tau} - y_u, \eta \rangle_2, \quad u \in S(\mathrm{P}_{\bar{\tau}}). \tag{6.52}$$

若还有 $\bar{u}$ 满足 (6.38), 则对所考虑的子列, 有目标函数的下述二阶展开

$$\mathrm{val}(\bar{\tau} + t_k\eta) = \mathrm{val}(\mathrm{P}_{\bar{\tau}}) + t_k\langle \bar{\tau} - y_{\bar{u}}, \eta \rangle_2 + \frac{1}{2}t_k^2\mathrm{val}(\mathrm{SP}) + o(t_k^2), \tag{6.53}$$

若 v 是 $(u_k - \bar{u})/t_k$ 在 $L_2(\Omega)$ 中的弱极限 (由 (i), 这样的极限是存在的), 则 v 是 $(u_k - \bar{u})/t_k$ 在 $L_2(\Omega)$ 中的强极限点, 且是 (SP) 的解.

证明　(i) 根据 (6.50), 由于 $N > 0$, 序列 $\{u_k\}$ 在 $L_s(\Omega)$ 中是有界的, 则至少有一弱极限点 $\bar{u}$. 因为 $(y, u, \tau) \mapsto J_\tau(y, u)$ 是从 $(L_2(\Omega))^3$ 到 $\mathbb{R}$ 的弱下半连续函数, 有

$$J_{\bar{\tau}}(\bar{y}, \bar{u}) \leqslant \liminf_k J_{\tau_k}(y_k, u_k) = \liminf_k \mathrm{val}(\tau_k).$$

另一方面, 由 $J_{\tau_k}(y_k, u_k) \leqslant J_{\tau_k}(\bar{y}, \bar{u})$, 有

$$\limsup_{k\to\infty} \mathrm{val}(\tau_k) \leqslant \limsup_{k\to\infty} J_{\tau_k}(\bar{y}, \bar{u}) \leqslant J_{\bar{\tau}}(\bar{y}, \bar{u}).$$

上述不等式实际上是等式, 因此有 $\bar{u} \in S(\mathrm{P}_{\bar{\tau}})$. 所以, $J_{\tau_k}(y_k, u_k) \to J_{\bar{\tau}}(\bar{y}, \bar{u})$, 因而 $\|u_k\|_2 \to \|\bar{u}\|_2$. 结果于 $L_2(\Omega)$ 中有 $u_k \to \bar{u}$.

令 $\bar{u} \in S(\mathrm{P}_{\bar{\tau}})$, 有

$$\mathrm{val}(\tau) \leqslant F(\bar{u}, \tau) = F(\bar{u}, \bar{\tau}) + D_\tau F(\bar{u}, \bar{\tau})(\tau - \bar{\tau}) + \frac{1}{2}\|\tau - \bar{\tau}\|_2^2.$$

令 $u_\tau \in S(\mathrm{P}_\tau)$. 用引理 6.30 之证明中的技术, 得到

$$\begin{aligned}\mathrm{val}(\tau) = F(u_\tau, \tau) = &F(\bar{u}, \bar{\tau}) + DF(\bar{u}, \bar{\tau})(u_\tau - \bar{u}, \tau - \bar{\tau}) \\ &+\frac{1}{2}D^2F(\bar{u}, \bar{\tau})((u_\tau - \bar{u}, \tau - \bar{\tau}), (u_\tau - \bar{u}, \tau - \bar{\tau})) \\ &+o(\|u_\tau - \bar{u}\|_2^2 + \|\tau - \bar{\tau}\|_2^2).\end{aligned}$$

结合这两个关系, 注意到一阶最优性条件 (6.23) 可推出 $DF(\bar{u}, \bar{\tau})(u_\tau - \bar{u}, \tau - \bar{\tau}) \geqslant D_\tau F(\bar{u}, \bar{\tau})(\tau - \bar{\tau})$, 得到

$$D^2F(\bar{u}, \bar{\tau})((u_\tau - \bar{u}, \tau - \bar{\tau}), (u_\tau - \bar{u}, \tau - \bar{\tau})) + o(\|u_\tau - \bar{u}\|_2^2 + \|\tau - \bar{\tau}\|_2^2) \leqslant \|\tau - \bar{\tau}\|_2^2.$$

考虑在 $L_s(\Omega)$ 中弱收敛到, 在 $L_2(\Omega)$ 中强收敛到 $\bar{u}$ 的子列 $u_k := u_{\tau_k}$. 若结论不真, 则有必要可抽取一子列, 有 $\|u_k - \bar{u}\|_2/\|\tau_k - \bar{\tau}\|_2 \to +\infty$. 由上述表达式可得

$$D_{u^2}^2 F(\bar{u}, \bar{\tau})(u_k - \bar{u}, u_k - \bar{u}) \leqslant o(\|u_k - \bar{u}\|_2^2).$$

令 $\bar{v}$ 是 $v_k := (u_k - \bar{u})/\|u_k - \bar{u}\|_2$ 在 $L_2(\Omega)$ 中的弱极限. 用上述不等式可得 $D_{u^2}^2 F(\bar{v}, \bar{v}) \leqslant 0, \bar{v} \in C(\bar{u})$, 因为 $D_{u^2}^2 F$ 是 Legendre 形式, $\bar{v} \neq 0$. 同二阶充分性最优条件(6.38) 矛盾.

(ii) 令 $u \in S(\mathrm{P}_{\bar{\tau}})$, 由于

$$\mathrm{val}(\tau_k) \leqslant F(u, \tau_k) = F(u, \bar{\tau}) + t_k D_\tau F(u, \bar{\tau})\eta + \frac{1}{2}t_k^2\|\eta\|_2^2,$$

得到

$$\begin{aligned}\mathrm{val}_+{}'(\bar{\tau}, \eta) :=&\limsup t_k^{-1}(\mathrm{val}(\bar{\tau} + t_k\eta) - \mathrm{val}(\bar{\tau})) \\ \leqslant&\inf\{D_\tau F(u, \bar{\tau})\eta : u \in S(\mathrm{P}_{\bar{\tau}})\}.\end{aligned} \tag{6.54}$$

另一方面, 对某一子序列 $u_k \in S(\mathrm{P}_{\tau_k})$, 它在 $L_s(\Omega)$ 中弱收敛到 $\bar{u}$. 由于 $D_\tau F(u, \bar{\tau}) = \bar{\tau} - y_u$, 有 $(u, \tau) \mapsto D_\tau F(u, \bar{\tau})$ 是 $L_2(\Omega) \times L_q(\Omega)$ 上的连续函数, 其中第一空间赋予弱拓扑. 所以,

$$\begin{aligned}\mathrm{val}(\tau_k) = F(u_k, \tau_k) &= F(u_k, \bar{\tau}) + t_k D_\tau F(u_k, \bar{\tau})\eta + o(t_k) \\ &\geqslant F(\bar{u}, \bar{\tau}) + t_k D_\tau F(\bar{u}, \bar{\tau})\eta + o(t_k).\end{aligned} \tag{6.55}$$

与 (6.54) 相结合, 得到 $\bar{u}$ 是 (LP) 的解.

现在设 $\bar{u}$ 满足 (6.38). 令 $v \in \mathcal{R}_K(\bar{u}) \cap D_u F(\bar{u}, \tau)^{\perp} \cap L_s(\Omega)$[①]. 将 F 展开, 有

$$\begin{aligned}\operatorname{val}(\bar{\tau}+t_k\eta) &\leqslant F(\bar{u}+t_k v, \bar{\tau}+t_k\eta) \\ &= \operatorname{val}(\bar{\tau}) + t_k D_\tau F(\bar{u},\bar{\tau})\eta + \frac{1}{2}t_k^2 D^2 F(\bar{u},\bar{\tau})((v,\eta),(v,\eta)) + o(t_k^2).\end{aligned}$$

由于 $D^2F(\bar{u},\bar{\tau})((v,\eta),(v,\eta))$ 是 (SP) 的目标函数, 用 s 多面性与 (6.34), 得到

$$\limsup_k \frac{\operatorname{val}(\bar{\tau}+t_k\eta) - \operatorname{val}(\bar{\tau}) - t_k D_\tau F(\bar{u},\bar{\tau})\eta}{\frac{1}{2}t_k^2} \leqslant \operatorname{val}(\mathrm{SP}). \tag{6.56}$$

另一方面, 由 (i), 若 $u_k \in S(\mathrm{P}_{\tau_k})$ 且 $u_k \to \bar{u}$, 序列 $v_k := (u_k - \bar{u})/t_k$ 至少有一弱极限点 v 在 $T_K(\bar{u})$ 中. 由

$$F(u_k, \tau_k) = F(\bar{u},\bar{\tau}) + t_k DF(\bar{u},\bar{\tau})(v_k,\eta) + \frac{1}{2}t_k^2 D^2F(\bar{u},\bar{\tau})(v_k,\eta)(v_k,\eta) + o(t_k^2)$$

与 D^2F 在 $L_2(\Omega)\times L_2(\Omega)$ 中的下半连续性, 可得 $v \in C(u)$ 与

$$\begin{aligned}F(u_k,\tau_k) &\geqslant F(\bar{u},\bar{\tau}) + t_k D_\tau F(\bar{u},\bar{\tau})\eta + \frac{1}{2}t_k^2 D^2F(\bar{u},\bar{\tau})(v,\eta)(v,\eta) + o(t_k^2) \\ &\geqslant F(\bar{u},\bar{\tau}) + t_k D_\tau F(\bar{u},\bar{\tau})\eta + \frac{1}{2}t_k^2 \operatorname{val}(\mathrm{SP}) + o(t_k^2).\end{aligned}$$

将此与 (6.56) 相结合, 即得目标函数的二阶展开与 $D^2F(\bar{u},\bar{\tau})(v_k,\eta)(v_k,\eta) \to \operatorname{val}(\mathrm{SP})$. 由于 $D^2_{u^2}F$ 是 Legendre 形式, 后一收敛关系可推出, 对于所考虑的子序列, $v_k \to v$ 于 $L_2(\Omega)$ 中是强收敛的. □

注 6.37 令 $n \leqslant 3$, 若 $\bar{u}$ 是 $S(\mathrm{P}_{\bar{\tau}})$ 中满足二阶增长条件 (6.36) 的局部解, 则在 $L_2(\Omega)$ 中存在 u[②]的闭邻域 $\mathcal{U}$ 满足局部问题:

$$(\mathrm{P}_{\tau,\mathcal{U}}) \qquad \min_u J_\tau(u, y_u) \quad \text{s.t.} \quad u \in \mathcal{U}, \quad u \in K \tag{6.57}$$

当 $\tau = \bar{\tau}$ 时有唯一的解 $\bar{u}$. 附加约束 $u \in \mathcal{U}$ 当 τ 接近于 $\bar{\tau}$ 时是不起作用的. 对上述的局部化问题可以做类似的扰动分析. 若还有问题(SP) 具有唯一解 $\bar{v}$, $u(t) \in S(\mathrm{P}_{\bar{\tau}+t\eta,\mathcal{U}})$, 则 $u(t) = \bar{u} + t\bar{v} + o(t)$, 即路径 $u(t)$ 具有右方向导数$\bar{v}$.

6.3.5 状态约束的最优控制问题

这一节用于讨论某些状态约束的最优控制问题. 这样的问题较约束仅加在控制上的问题要难得多. 因此, 讨论限定在 $n \leqslant 3$ 的情况. 考虑最优控制问题:

$$(\mathrm{P}_\tau^\delta) \qquad \min_u J_\tau(u, y_u) \quad \text{s.t.} \quad u \in K, \quad y \geqslant \delta \ \text{在}\Omega\text{上}, \tag{6.58}$$

① 应为 $D_u F(\bar{u}, \bar{\tau})^{\perp}$.

② 应为 $\bar{u}$.

这里 $y \geqslant \delta$ 意味着对所有的 $\omega \in \Omega, y(\omega) \geqslant \delta$, δ 是负的实数值, 从而状态约束与边界条件是相容的. 由于状态是连续的 (因为 $n \leqslant 3$), 准时状态约束是有意义的. 最优解的存在性可用 6.3.1 节中所采用的类似方式证明.

为了叙述一阶最优性系统, 需要讨论右端有测度的协态方程:

$$\begin{cases} -\Delta p + \phi'(y_u)p = \mu, & \text{在 } \Omega \text{ 内}, \\ p = 0, & \text{在 } \partial\,\Omega \text{ 上}, \end{cases} \tag{6.59}$$

其中 $\mu \in C_0(\Omega)^*$, 即 μ 是在 Ω 上支撑的有限符号的 Radon 测度. 以分布定义解释 (6.59), 即 $p \in W_0^{1,1}(\Omega)$ 是 (6.59) 解的充要条件是

$$\int_\Omega (-\Delta z + \phi'(y_u)z)(\omega)p(\omega)\mathrm{d}\omega = \int_\Omega z(\omega)\mathrm{d}\mu(\omega), \quad \forall z \in \mathcal{D}(\Omega). \tag{6.60}$$

由于 $\mathcal{D}(\Omega)$ 是 $C_0(\Omega)$ 的稠密子集, 这意味着分布 $-\Delta p$ 到 $C_0(\Omega)$ 上具有唯一的扩展, 等于 $\mu - \phi'(y_u)p$, 仍用 $-\Delta p$ 来记. 所以, 等式 $-\Delta p + \phi'(y_u)p = \mu$ 在 $C_0(\Omega)^*$ 内也是有效的.

引理 6.38 *对所有的 $t \in [1, n/(n-1))$, 方程 (6.59) 有唯一的解 $p \in W_0^{1,t}(\Omega)$, 且对这样的 t, 存在 $c_t > 0$ 满足*

$$\|p\|_{1,t} \leqslant c_t \|\mu\|_{C_0(\Omega)}. \tag{6.61}$$

证明 给定 $f \in L_2(\Omega)$, 线性化状态方程:

$$\begin{cases} -\Delta z + \phi'(y_u)z = f, & \text{在 } \Omega \text{ 内}, \\ z = 0, & \text{在 } \partial\,\Omega \text{ 上} \end{cases} \tag{6.62}$$

有唯一的解 $z \in Y$, 其中 $Y := H^2(\Omega) \cap H_0^1(\Omega)$. 因为 $n \leqslant 3$, Y 被连续地嵌入于 $C_0(\Omega)$ 中, 映射 $f \mapsto Af := z$ 是 $L_2(\Omega) \to C_0(\Omega)$ 的线性连续映射. 伴随算子 $A^* : C_0(\Omega)^* \to L_2(\Omega)$ 满足, 若 $p = A^*\mu, \mu \in C_0(\Omega)^*$, 则 p 可被下述关系刻画:

$$\langle f, p \rangle_{L_2(\Omega)} = \langle \mu, Af \rangle_{C_0(\Omega)^*, C_0\Omega)}, \quad \forall f \in L_2(\Omega). \tag{6.63}$$

另一方面, 令 t, t' 满足 $t' > n, 1/t + 1/t' = 1$. 根据文献 [160] 中的定理1.4, 只要 $f \in W^{-1,t'}(\Omega)$, (6.62) 有唯一的解 $z \in W_0^{1,t'}(\Omega)$. 因为 $t' > n$, 由引理 6.4, $W_0^{1,t'}(\Omega) \subset C_0(\Omega)$ 具有连续的嵌入. 所以, $f \mapsto z$ 是由 $W^{-1,t'}(\Omega)$ 到 $C_0(\Omega)$ 的线性连续映射, 用 $\bar{A}$ 来记. 伴随算子$\bar{A}^* : C_0(\Omega)^* \to W_0^{1,t}(\Omega)$ 满足 $\hat{p} = \bar{A}^*\mu$, $\mu \in C_0(\Omega)^*$ 被下述关系刻画:

$$\langle \hat{p}, f \rangle_{W_0^{1,t}(\Omega), W^{-1,t'}(\Omega)} = \langle \mu, \bar{A}f \rangle_{C_0(\Omega)^*, C_0\Omega)}, \quad \forall f \in W^{-1,t'}(\Omega). \tag{6.64}$$

因为 $W_0^{1,t}(\Omega)$ 是 $L_t(\Omega)$ 的稠密子集[1], 根据引理 6.3, $L_{t'}(\Omega)$ 是 $W^{-1,t'}(\Omega)$ 的稠密子集. 因为 $L_{t'}(\Omega)$ 也是 $L_2(\Omega)$ 的稠密子集, 方程 (6.63) 与 (6.64) 具有相同的解, 即 $p=\hat{p}$. 结果得证. □

定理 6.39　若 $\bar{u}$ 是 (P_τ^δ) 与状态 $\bar{y}$ 相联系的局部解, 对所有的 $t \in [1, n/(n-1))$, 存在 $\alpha \geqslant 0, \bar{p} \in W_0^{1,t}(\Omega)$ 及 $\mu \in C_0(\Omega)_-^*$, 满足 $\alpha + \|\mu\|_{C_0(\Omega)^*} > 0$ 且

$$\begin{cases} -\Delta\bar{p} + \phi'(\bar{y})\bar{p} = \alpha(\bar{y} - \tau) + \mu, & \text{在 } \Omega \text{ 内}, \\ \bar{p} = 0, & \text{在 } \partial\Omega \text{ 上}, \end{cases} \tag{6.65}$$

$$\operatorname{supp}(\mu) \subset \{\omega \in \Omega : \bar{y}(\omega) = \delta\}, \tag{6.66}$$

$$0 \in \alpha N\bar{u} + \bar{p} + N_K(\bar{u}). \tag{6.67}$$

证明　根据命题 3.16, 因为集合 $\{z \in C_0(\Omega) : z \geqslant \delta\}$ 具有非空内部, 存在广义的Lagrange 乘子. 将此与引理 6.18 与 6.38 相结合, 得到上述最优性系统. □

只要上述定理的结论成立, 就称 (α, p, μ) 是与 $\bar{u}$ 相联系的广义的Lagrange 乘子(没有显式地对应 $N_K(\bar{u})$ 的元素). 现在转向充分性最优条件的研究, 这一表述用到 Lagrange 函数的下述表示:

$$L(y, u, p, \mu) := J_\tau(y, u) - \int_\Omega p(\omega)(-\Delta y + \phi(y) - u)(\omega)\mathrm{d}\omega + \int_\Omega y(\omega)\mathrm{d}\mu(\omega).$$

定理 6.40　设 $\bar{u}$ 是 (P) 的可行点, $\bar{y}$ 是相应的状态, 满足对所有的非零临界方向 v, z_v 为 $f=v$ 时线性化状态方程 (6.62) 的解, 存在 Lagrange 乘子(p, μ), 满足

$$D^2_{(u,y)^2}L(\bar{u}, \bar{y}, p, \mu)((v, z_v), (v, z_v)) > 0.$$

则 $\bar{u}$ 是 (P) 的满足二阶增长条件的局部解.

证明　这是定理 3.63、命题 3.74 与 3.77 的标准应用. □

6.3.6　病态系统的最优控制

现在转向讨论具有病态方程的问题, 即

$$\begin{cases} -\Delta y - y^3 = u, & \text{在 } \Omega \text{ 内}, \\ y = 0, & \text{在 } \partial\Omega \text{ 上}. \end{cases} \tag{6.68}$$

这里设 $n \leqslant 3$. (6.68) 的解意味着某一 $y \in H_0^1(\Omega)$ 满足分布意义下的第一个方程. 因为当 $n \leqslant 3$ 时 $H_0^1(\Omega) \subset L_6(\Omega)$, 有 $-\Delta y \in L_2(\Omega)$, 因此 $y \in H^2(\Omega)$ 是 (6.68) 的强解.

以下验证这一方程可能有几个解. 当 $u=0$ 时, 有平凡解 $y=0$, 且若 y 是解, 则 $-y$ 亦是解. 由下面的引理得, 若 $u=0$, 则 (6.68) 至少有三个解.

引理 6.41 *若 $u=0$, 则 (6.68) 有非零解.*

证明 考虑非凸的变分问题:

$$\min_z \frac{1}{2}\int_\Omega |\Delta z(\omega)|^2 \mathrm{d}\omega, \quad \frac{1}{4}\int_\Omega z(\omega)^4 \mathrm{d}\omega \geqslant 1, \quad z \in H_0^1(\Omega). \tag{6.69}$$

以下证明这一问题有解. 根据 Poincaré不等式, 极小化序列在 $H_0^1(\Omega)$ 中是有界的. 由 Sobolev 包含关系, 在 $H_0^1(\Omega)$ 中弱收敛的子序列在 $L_4(\Omega)$ 中是强收敛的. 因为 (6.69) 中的目标函数是弱下半连续的, 这一子序列的极限是 (6.69) 的解.

下一步是要证 (6.68) 的非零解的存在性. 令 $\bar{z}$ 是 (6.69) 的解. 由于问题的目标函数与约束函数是连续可微的, 约束的导数, 即 z^3 在任何可行点处均是非零的, 因此存在 Lagrange 乘子 $\lambda \geqslant 0$ 满足在对偶空间 $H^{-1}(\Omega)$ 中最优性条件 $-\Delta\bar{z}-\lambda\bar{z}^3=0$ 成立. 还有 $\lambda \neq 0$, 因为否则 $\bar{z}=0$ 是不可行的. 有 $y:=\lambda^{1/2}\bar{z}$ 是 (6.68) 的非零解. □

注 6.42 若 z 是 (6.69) 的解, 则根据命题 6.45, $|z|$ 亦是解. 从而 (6.68) 具有非零的非负的解 $\bar{y}$.

考虑目标函数:

$$J(y,u) := \frac{1}{6}\int_\Omega (y(\omega)-\tau(\omega))^6 \mathrm{d}\omega + \frac{1}{2}N\int_\Omega u(\omega)^2 \mathrm{d}\omega,$$

其最优控制问题是

$$\text{(P)} \qquad \min_{y,u} J(y,u) \text{ s.t. } (6.68), \quad u \in K, \quad y \geqslant \delta, \tag{6.70}$$

其中 $\tau \in L_6(\Omega)$, δ 是负实数. 为了验证最优控制的存在性, 取非二次的目标函数.

定理 6.43 *(i) 问题(P) 当其可行时 (至少) 有解存在; (ii) 若 $\bar{u}$ 是 (P) 与状态 $\bar{y}$ 相联系的局部解, 则对所有的 $t \in [1, n/(n-1))$, 存在 $\alpha \geqslant 0, \bar{p} \in W_0^{1,t}(\Omega)$ 与 $\mu \in C_0(\Omega)_-^*$, 满足 $\alpha + \|\bar{p}\|_{L_2(\Omega)} > 0$, 且*

$$\begin{cases} -\Delta\bar{p} - 3\bar{y}^2\bar{p} = \alpha(\bar{y}-\tau)^5 + \mu, & \text{在 } \Omega \text{ 内}, \\ \bar{p} = 0, & \text{在 } \partial\Omega \text{ 上}, \end{cases} \tag{6.71}$$

$$\operatorname{supp}(\mu) \subset \{\omega \in \Omega : \bar{y}(\omega) = \delta\}, \tag{6.72}$$

$$0 \in \alpha N\bar{u} + \bar{p} + N_K(\bar{u}). \tag{6.73}$$

证明 (i) 只需注意到, 由目标函数可推出极小化序列(y_k, u_k) 在 $L_6(\Omega)\times L_2(\Omega)$ 内是有界的, 而 $-\Delta y_k = y_k^3 + u_k$ 在 $L_2(\Omega)$ 中是有界的, 因此 y_k 在 $H^2(\Omega)$ 内是有界的, 类似于定理 6.16 的证明, 可以建立最优控制的存在性.

(ii) 证明广义Lagrange 乘子的存在性, 通过对状态方程为

$$\begin{cases}-\Delta y = u + v, & \text{在 } \Omega \text{ 内},\\ y = 0, & \text{在 } \partial\Omega \text{ 上},\end{cases} \tag{6.74}$$

以 u 与 v 为控制, 目标函数为

$$J_\varepsilon(y,u,v) := J(y,u) + \frac{1}{\varepsilon}\int_\Omega (y(\omega)^3 - v(\omega))^2 \mathrm{d}\omega + \frac{1}{2}\int_\Omega (y(\omega) - \bar{y}(\omega))^2 \mathrm{d}\omega$$

的惩罚问题的最优性系统取极限而得到. $J_\varepsilon(y,u,v)$ 的第二项对于小的 ε, 强迫 v 接近于 y^3, 从而使原始的状态的方程被近似地满足, 最后一项确保惩罚问题的解收敛到某一具体的解 $(\bar{y},\bar{u})$. 最优控制问题是

$$(\mathrm{P}_\varepsilon) \qquad \min_{y,u,v} J_\varepsilon(y,u,v) \quad \text{s.t.} \quad u \in K, \quad y \geqslant \delta. \tag{6.75}$$

给定 $\varepsilon > 0$, 通过对问题(P_ε) 的极小化序列取极限, 容易看到这一问题 (至少) 有一解 $(y_\varepsilon, u_\varepsilon, v_\varepsilon)$, 其中 $y_\varepsilon \in H^2(\Omega) \cap H_0^1(\Omega), u_\varepsilon, v_\varepsilon$ 在 $L_2(\Omega)$ 中. 因为

$$J_\varepsilon(y_\varepsilon, u_\varepsilon, v_\varepsilon) \leqslant J(\bar{y},\bar{u}), \tag{6.76}$$

有在 $L_6(\Omega)\times L_2(\Omega)$ 中 $(y_\varepsilon, u_\varepsilon)$ 的一致估计, 从而有 $L_2(\Omega)$ 中 v_ε 的一致估计, 最后有在 $H^2(\Omega)$ 中 y_ε 的一致估计.

令 $(\tilde{y},\tilde{u},\tilde{v})$ 是 $(y_\varepsilon, u_\varepsilon, v_\varepsilon)$ 在 $H^2(\Omega)\cap H_0^1(\Omega)$ 与 $L_2(\Omega)$ 及 $L_2(\Omega)$ 中的弱极限. 注意到, $y_\varepsilon \to \tilde{y}$ 在 $C_0(\Omega)$ 内是强收敛的. 由 (6.76), 有 $\int_\Omega (y_\varepsilon(\omega)^3 - v_\varepsilon(\omega))^2 \mathrm{d}\omega \to 0$. 因为在 $C_0(\Omega)$ 中 $y_\varepsilon \to \tilde{y}$, 所以在 $L_2(\Omega)$ 中 v_ε 强收敛到 $\tilde{v} = \tilde{y}^3$, 因此 $(\tilde{y},\tilde{u})$ 是 (6.68) 的解. 另一方面, 目标函数 J 是 $H^2(\Omega)\times L_2(\Omega) \to \mathbb{R}$ 的凸的连续的函数, 因此是弱下半连续的, 有

$$J(\tilde{y},\tilde{u}) + \frac{1}{2}\int_\Omega (y(\omega) - \bar{y}(\omega))^2 \mathrm{d}\omega \leqslant \liminf J(y_\varepsilon; u_\varepsilon) \leqslant J(\bar{y},\bar{u}).$$

证得 $J(\tilde{y},\tilde{u}) = \lim J(y_\varepsilon, u_\varepsilon)$, 从而 $(\tilde{y},\tilde{u})$ 等于 $(\bar{y},\bar{u})$. 已经证得 y_ε 与 u_ε 在 $H^2(\Omega)$ 与 $L_2(\Omega)$ 中弱收敛到 $\bar{y}$ 与 $\bar{u}$, 但因为 J 是这两个空间两个连续的凸函数之和, 实际上有 $\int_\Omega u_\varepsilon(\omega)^2 \mathrm{d}\omega \to \int_\Omega \bar{u}(\omega)^2 \mathrm{d}\omega$. 因此在 $L_2(\Omega)$ 中有 $u_\varepsilon \to \bar{u}$. 因为在相同的空间有 $v_\varepsilon \to \bar{y}^3$, 由 (6.74) 得, 在 $H^2(\Omega)\cap H_0^1(\Omega)$ 中有 $y_\varepsilon \to \bar{y}$.

由于问题 (P_ε) 的状态方程是适定的, 根据上一节的推证可得, 对于所有的 $t \in [1, n/(n-1))$, 存在 $\alpha_\varepsilon \in \mathbb{R}_+$, $p_\varepsilon \in W_0^{1,t}(\Omega)$ 与 $\mu_\varepsilon \in C_0(\Omega)^*_-$, 它们不全为 0, 满足

$$\begin{aligned}&-\Delta p_\varepsilon = \alpha_\varepsilon (y_\varepsilon - \tau)^5 + 6\alpha_\varepsilon \varepsilon^{-1} y_\varepsilon^2 (y_\varepsilon^3 - v_\varepsilon) + \alpha_\varepsilon (y_\varepsilon - \bar{y}) + \mu_\varepsilon, \quad \text{在 } \Omega \text{ 内},\\ &\operatorname{supp}(\mu_\varepsilon) \subset \{\omega \in \Omega : y_\varepsilon(\omega) = \delta\},\\ &0 \in \alpha_\varepsilon N u_\varepsilon + p_\varepsilon + N_K(u_\varepsilon), \quad p_\varepsilon + 2\alpha_\varepsilon \varepsilon^{-1}(v_\varepsilon - y_\varepsilon^3) = 0.\end{aligned}$$

因为上述系统的解集是锥, 不妨设 $\alpha_\varepsilon + \|p_\varepsilon\|_{L_2(\Omega)} + \|\mu\|_{C_0(\Omega)^*} = 1$. 将最后一个方程中的 v_ε 消掉, 用到法锥的定义, 得到等价系统

$$
\begin{aligned}
&-\Delta p_\varepsilon - 3y_\varepsilon^2 p_\varepsilon = \alpha_\varepsilon(y_\varepsilon - \tau)^5 + \alpha_\varepsilon(y_\varepsilon - \bar{y}) + \mu_\varepsilon, \quad \text{在 } \Omega \text{ 内},\\
&\int_\Omega (z(\omega) - y_\varepsilon(\omega))\mathrm{d}\mu_\varepsilon(\omega) \leqslant 0, \quad \forall z \in C_0(\Omega), \quad z \geqslant \delta,\\
&\int_\Omega (\alpha_\varepsilon N u_\varepsilon(\omega) + p_\varepsilon(\omega)(v(\omega) - u_\varepsilon(\omega))\mathrm{d}\omega \geqslant 0, \quad \forall v \in K.
\end{aligned}
$$

因为 p_ε 在 $L_2(\Omega)$ 中有界, 根据引理 6.38(这里 $\phi = 0$), 可推出对所有的 $t \in [1, n/(n-1))$, p_ε 在 $W_0^{1,t}(\Omega)$ 内有界. 因为只要 $t_1 \leqslant t_2$, 就有 $W_0^{1,t_1}(\Omega) \subset W_0^{1,t_2}(\Omega)$, 对某一序列 $\varepsilon_k \to 0$, 得 p_{ε_k} 弱收敛到 $W_0^{1,t}(\Omega)$ 中的某元素 $\bar{p}$. 因此 (因 $n \leqslant 3$, 可由 Sobolev 包含关系) p_{ε_k} 在 $L_2(\Omega)$ 内收敛到 $\bar{p}$, μ_{ε_k} 弱* 收敛到 $C_0(\Omega)^*$ 中的某一元素 μ(因为定理2.28可应用到这里), 且 $\alpha_{\varepsilon_k} \to \alpha \geqslant 0$. 将上述的关系取极限, 可得到 (6.71) ~ (6.73).

剩下需证 α 与 $\bar{p}$ 不全为零. 用反证法. 若 α 与 $\bar{p}$ 均等于零, 且 $\alpha_{\varepsilon_k} \to 0$, 在 $L_2(\Omega)$ 中有 $p_\varepsilon \to 0$, 因为它在此空间内是强收敛的, 因此有 $\|\mu\|_{C_0(\Omega)^*} \to 1$, 而由 (6.71) 得 $\mu = 0$.

另一方面, 由于 $\delta < 0$, 有

$$
\int_\Omega y_\varepsilon(\omega)\mathrm{d}\mu_\varepsilon(\omega) \geqslant \sup\left\{\int_\Omega z(\omega)\mathrm{d}\mu_\varepsilon(\omega) : z \in C_0(\Omega), \max(|z|) \leqslant \delta\right\},
$$

因此, 在 $C_0(\Omega)$ 中有 $y_\varepsilon \to \bar{y}$,

$$
\int_\Omega \bar{y}(\omega)\mathrm{d}\mu(\omega) = \lim_{k\to\infty}\int_\Omega y_{\varepsilon_k}(\omega)\mathrm{d}\mu_{\varepsilon_k}(\omega) \geqslant \lim_{k\to\infty} \delta\|\mu_{\varepsilon_k}\|_{C_0(\Omega)^*} = \delta.
$$

这与 $\mu = 0$ 是矛盾的. □

6.4 障 碍 问 题

这一节讨论在偏微分方程的框架下变分不等式的最简单的例子, 即著名的障碍问题. 对这一问题有若干处理方法, 其中某些方法在本书最后的注记中给出评述. 这一节仅仅给出本书中某些抽象结果在此问题中的应用.

6.4.1 问题的表述

问题可以表示为

$$
\min\left\{\Phi(y) := \frac{1}{2}\int_\Omega |\nabla y(\omega)|^2\mathrm{d}\omega - \int_\Omega u(\omega)y(\omega)\mathrm{d}\omega\right\}, \quad y \geqslant 0 \text{ a.e.} \quad \text{在 } \Omega \text{ 上}.
$$

视 $u \in L_2(\Omega)$ 为控制变量, y 视为相联系的状态变量. 因为 Φ 是凸函数 (目标函数), 它的极小由相应的最优性系统刻画:

$$\int_\Omega \nabla y(\omega) \cdot (\nabla z(\omega) - \nabla y(\omega))\mathrm{d}\omega \geqslant \int_\Omega (z(\omega) - y(\omega))u(\omega)\mathrm{d}\omega,$$
$$\forall z \in H_0^1(\Omega)_+, \qquad y \in H_0^1(\Omega)_+. \tag{6.77}$$

以下验证障碍问题是适定的.

定理 6.44 对任何 $u \in L_2(\Omega)$, 障碍问题具有唯一解 $y_u \in H^2(\Omega) \cap H_0^1(\Omega)$, 并且, 这是从 $w - L_2(\Omega)$ 到 $w - H^2(\Omega)$ 的连续映射.

证明 上述极小化问题的目标函数是连续的, 根据 Poincaré引理, 它在 $H_0^1(\Omega)$ 上是强凸的. 所以, 唯一极小点的存在性是引理 2.33 的一个结果.

用惩罚形式的推证转向 $H^2(\Omega)$ 正则性的讨论. 设 β 是 $\mathbb{R} \to \mathbb{R}$ 的凸的 Lipschitz 连续的 C^∞ 光滑函数, 满足: 若 $\sigma \geqslant 0$, 则 $\beta(\sigma) = 0$, 若 $\sigma < 0$, 则 $0 < \beta(\sigma) < -\sigma$. 考虑问题:

$$\min \left\{ \Phi_\varepsilon(y) := \frac{1}{2}\int_\Omega |\nabla y(\omega)|^2\mathrm{d}\omega - \int_\Omega u(\omega)y(\omega)\mathrm{d}\omega + \frac{1}{\varepsilon}\int_\Omega \beta(y(\omega))\mathrm{d}\omega \right\}.$$

由于 Φ_ε 在 $H_0^1(\Omega)$ 中的有界子集上是强凸的有界的, 根据命题 2.108 , 它是连续的. 对任何 $\varepsilon > 0$, 由引理 2.33, 上述极小化问题有唯一的解 $y_\varepsilon \in H^2(\Omega) \cap H_0^1(\Omega)$, 由方程

$$-\Delta y_\varepsilon + \beta'(y_\varepsilon) = u, \quad 在\ \Omega\ 上$$

刻画, 乘以 $-\Delta y_\varepsilon$, 在 Ω 上取积分, 得到①

$$\|\Delta y_\varepsilon\|_2^2 + \int_\Omega \beta''(y_\varepsilon(\omega))|\nabla y(\omega)|^2\mathrm{d}\omega \leqslant \|\Delta y_\varepsilon\|_2 \cdot \|u\|_2.$$

由于第二项积分是非负的, 有 $\|\Delta y_\varepsilon\|_2 \leqslant \|u\|_2$. 当 $\varepsilon \downarrow 0$ 时, 由于 y_ε 在 $H_2(\Omega)$ 中是有界的, 则它在 $H^2(\Omega)$ 中至少有一弱极限点 $\bar{y}$. 因为 $y_u \geqslant 0$, 有 $\Phi_\varepsilon(y_\varepsilon) \leqslant \Phi_\varepsilon(y_u) \leqslant \Phi(y_u)$, 因而 $\Phi(\bar{y}) \leqslant \liminf_{\varepsilon\downarrow 0} \Phi(y_\varepsilon) \leqslant \Phi(y_u)$. 另一方面, 由 $\Phi(y_\varepsilon) \leqslant \Phi(y_u)$ 可得

$$\int_\Omega \beta(y_\varepsilon(\omega))\mathrm{d}\omega \leqslant \varepsilon(\Phi(y_u) - \Phi(y_\varepsilon)).$$

因为 $\Phi(y_\varepsilon)$ 是有界的, $y \mapsto \int_\Omega \beta(y(\omega))\mathrm{d}\omega$ 是 $H_0^1(\Omega)$ 上的凸的连续函数, 因而是弱下半连续的 (见例 2.30), 可推出

$$\int_\Omega \beta(\bar{y}(\omega))\mathrm{d}\omega \leqslant \liminf_{\varepsilon\downarrow 0} \int_\Omega \beta(y_\varepsilon(\omega))\mathrm{d}\omega \leqslant 0.$$

---

① 原文中下式第一项为 $\|\Delta y_\varepsilon\|_2$.

于是得到 $\bar{y} \geqslant 0$. 由 $\Phi(\bar{y}) \leqslant \Phi(y_u)$, 可得 $\bar{y} = y_u, \|\Delta y_u\|_2 \leqslant \|u\|_2$.

最后, 如果在 $L_2(\Omega)$ 中 $u_n \xrightarrow{w} u$, 则 $y_n := y_{u_n}$ 满足 Δy_n 在 $L_2(\Omega)$ 中是有界的, 因而存在 $y \in H^2(\Omega)$ 满足对某一子序列, 在 $H^2(\Omega)$ 中有 $y_n \xrightarrow{w} y$, 且在 $H_0^1(\Omega)$ 中有 $y_n \to y$. 因此, 可以在最优条件中取极限, 得

$$\int_\Omega \nabla y_n \cdot \nabla(z - y_n)(\omega)\mathrm{d}\omega \geqslant \int_\Omega u_n(\omega)(z - y_n)(\omega)\mathrm{d}\omega, \quad \forall z \in H_0^1(\Omega)_+,$$

这推得对所有的 $z \in H_0^1(\Omega)_+$, $D\Phi(y)(z - y) \geqslant 0$. 这意味着 Φ 在 $H_0^1(\Omega)$ 上于 y 处取得极小值. 因为 Φ 是强凸的, 可推出 $y = y_u$. □

6.4.2 多面性

现在验证 $H_0^1(\Omega)$ 是 Hilbert 格. 更一般地, 对于 $s \geqslant 2$, 证明 $W_0^{1,s}(\Omega)$ 是 Banach 格. 用 $1_{\{u>0\}}$ 记集合 $u > 0$ 的指示函数, 即几乎处处定义的 $\Omega \to \mathbb{R}$ 函数, 若 $u(\omega) > 0$, 则 $1_{\{u>0\}}(\omega) = 1$, 否则取 0 值.

命题 6.45 *令 $u \in W_0^{1,s}(\Omega)$, 其中 $s \in [2, \infty\}$, 则函数 $u_+(\omega) = \max(u(\omega), 0)$ 满足 $u_+ \in W_0^{1,s}(\Omega)$, 并且 $u \mapsto u_+$ 是从 $W_0^{1,s}(\Omega)$ 到其自身的连续映射, 且*

$$\nabla u_+ = 1_{u>0}\nabla u, \quad \nabla u_+ \cdot \nabla u = |\nabla u_+|^2, \quad \text{a.e. 在 } \Omega \text{ 内}. \tag{6.78}$$

证明 存在连续可微的 Lipschitz 连续函数 $\phi: \mathbb{R} \to \mathbb{R}$ 满足 $0 \leqslant \phi' \leqslant 3$, 且

$$\phi(\sigma) = 0, \quad 若\ \sigma \leqslant \frac{1}{2}, \quad \phi(\sigma) \in [0, 1], \quad 若\ \sigma \in \left(\frac{1}{2}, 1\right), \quad \phi(\sigma) = \sigma, \quad 若\ \sigma \geqslant 1.$$

给定 $\varepsilon > 0$, 置 $\phi_\varepsilon(\sigma) := \varepsilon\phi(\varepsilon^{-1}\sigma)$. 由引理 6.11, 有 $\phi_\varepsilon(u) \in W_0^{1,s}(\Omega)$, 满足 $\nabla\phi_\varepsilon(u) = \phi_\varepsilon{}'(u)\nabla u = \phi'(\varepsilon^{-1}u)\nabla u$. 并且几乎处处地成立

$$|\nabla\phi_\varepsilon(u) - 1_{u>0}\nabla u| = |\phi'_\varepsilon(u) - 1_{u>0}| \cdot |\nabla u| \leqslant 2|1_{0<u<\varepsilon}| \cdot |\nabla u|.$$

所以, 根据 Lebesgue 控制收敛定理, 有 $\nabla\phi_\varepsilon(u) \to 1_{u>0}\nabla u$ 在 $L_s(\Omega)$ 中成立. 由 $\phi_\varepsilon(u) \to u_+$ a.e. 得到, 在 $W_0^{1,s}(\Omega)$ 中有 $\phi_\varepsilon(u) \to u_+$ 且 $\nabla u_+ = 1_{u>0}\nabla u$, 因此 $|\nabla u_+|^2 = 1_{u>0}|\nabla u|^2 = \nabla u_+ \cdot \nabla u$.

最后验证 $u \mapsto u_+$ 在 $W_0^{1,s}(\Omega)$ 中是连续的. 令 u 与 v 属于 $W_0^{1,s}(\Omega)$,

$$|\nabla u_+ - \nabla v_+| = |1_{u>0}\nabla u - 1_{v>0}\nabla v| \leqslant |1_{v>0}(\nabla u - \nabla v)| + |(1_{v>0} - 1_{u>0})\nabla u|.$$

根据 Lebesgue 控制定理, 在 $W_0^{1,s}(\Omega)$ 中当 $v \to u$ 时, 上式右端在 $L_2(\Omega)$ 中趋于 0. 结论得证. □

推论 6.46 *若 $s \in [2, \infty]$, 则 $W_0^{1,s}(\Omega)$ 的正锥是多面性的.*

证明 结合上述命题与定理 3.58 即可得到. □

由于 $H_0^1(\Omega)_+$ 是多面性的, 可用 4.7 节中的抽象灵敏度分析, 这涉及 $H_0^1(\Omega)_+$ 的切锥的表达式. 为了给出这一切锥的表达式, 需要容量理论的某些知识.

6.4.3 基本容量理论

除非声明, 这一节, 范数是 $H_0^1(\Omega)$ 的范数, 即 $\|\cdot\|_{1,2,0}$, 或相联系的对偶范数. 类似地, 对偶积是 $H_0^1(\Omega)$ 与其对偶间的对偶性积. 由定义可知, 集合 A 的邻域是这样一个集合, 它是 A 中每一元素的邻域. Borel 集在例 2.37 中定义.

定义 6.47 令 A 是 Ω 的 Borel 子集, $\alpha \in \mathbb{R}$.

(i) 在 $H_0^1(\Omega)$ 的意义下, 称 $u \in H_0^1(\Omega)$ 在 A 上满足 $u \geqslant \alpha$, 如果存在于 $H_0^1(\Omega)$ 中 $u_n \to u$ 的序列满足于 A 的邻域上几乎处处有 $u_n \geqslant \alpha$.

(ii) 在 $H_0^1(\Omega)$ 的意义下, A 的容量定义为

$$\operatorname{cap}(A) := \inf\{\|u\|^2 : \text{在} H_0^1(\Omega) \text{中的意义下于} A \text{上有} u \geqslant 1\}.$$

(iii) 称可测函数 $f : \Omega \to \mathbb{R}$ 是拟连续的, 如果在 Ω 上存在非增的开集序列 $\{\Omega_n\}$ 满足 f 在 $\Omega \backslash \Omega_n$ 上是连续的且 $\operatorname{cap}(\Omega_n) \to 0$.

例如, 考虑集合 $A = \{\omega_0\}$, 其中 $\omega_0 \in \Omega$, 则不难验证 A 具有零容量当且仅当 $n > 1$.

函数 f 几乎处处定义于 Ω 上, 如 $L_s(\Omega)$ 的元素与 Sobolev 空间$W^{m,s}(\Omega)$ 的元素, 实际上是定义于 Ω 上的函数的等价类, 其等价关系定义为: 如果于 Ω 上几乎处处有 $f(\omega) = g(\omega)$, 则有等价关系 $f \sim g$. 称 f 的等价类的元素为 f 的代表元素. 可能发生 f 的代表元素是连续的情况, 这时称 f 本身是连续的 (很清楚, 并不是 f 的每一代表元素均是连续的). 对于其他形式的连续性, 也可采用类似的约定, 如拟连续性. 为强调代表元函数不是几乎处处定义的函数, 我们称之为处处定义的函数.

若 f 在 Ω 上是处处定义的函数, 用 $\{f < 0\}$ 记集合 $\{\omega \in \Omega : f(\omega) < 0\}$.

引理 6.48 令 A_n 是 Ω 中的一列 Borel 子集, 则

$$\operatorname{cap}(\bigcup_n A_n) \leqslant \sum_n \operatorname{cap}(A_n). \tag{6.79}$$

证明 先证明, 若 u 与 v 属于 $H_0^1(\Omega)$, 则

$$\|\max(u,v)\|^2 \leqslant \|u\|^2 + \|v\|^2. \tag{6.80}$$

事实上, 令 $w \in H_0^1(\Omega)$. 根据命题 6.45, 有 $\langle w_-, w_+\rangle = 0$. 因为 $w = w_+ + w_-$, 有 $\langle w, w_+\rangle = \|w_+\|^2$. 因此

$$\begin{aligned}\|u + w_+\|^2 &= \|u\|^2 + 2\langle u, w_+\rangle + \|w_+\|^2\\ &= \|u\|^2 + 2\langle u + w, w_+\rangle - \|w_+\|^2\\ &= \|u\|^2 + \|u + w\|^2 - \|w + u - w_+\|^2\\ &\leqslant \|u\|^2 + \|u + w\|^2.\end{aligned}$$

取 $w = v - u$, 由 $u + w_+ = \max(u, v)$, 得到 (6.80).

其次, 观察到, 若 $\sum_n \operatorname{cap}(A_n) = \infty$, 则结论显然成立. 否则, 固定 $\varepsilon > 0$, 令 $u_n \in H_0^1(\Omega)$ 满足于 A_n 的邻域上有 $u_n \geqslant 1$ 且 $\|u_n\|^2 \leqslant \operatorname{cap}(A_n) + 2^{-n-1}\varepsilon$. 置

$$\hat{u}_n := \sup_{0 \leqslant i \leqslant n} u_i, \quad \bar{u} := \sup_n u_n.$$

因为 $H_0^1(\Omega)$ 是 Hilbert 格, 有 $\hat{u}_n \in H_0^1(\Omega)$, 且由 (6.80) 得

$$\|\hat{u}_n\|^2 \leqslant \sum_{i=0}^{n} \|u_i\|^2 \leqslant \sum_{n=0}^{n} \operatorname{cap}(A_i) + \varepsilon.$$

序列 $\{\hat{u}_n\}$ 在 $H_0^1(\Omega)$ 中是有界的且非递减的, 因此在 Ω 上几乎处处收敛于 $\bar{u}$. 所以, 它于 $L_2(\Omega)$ 中强收敛到 $\bar{u}$, 于 $H_0^1(\Omega)$ 中弱收敛到 $\bar{u}$. 结果,

$$\|\bar{u}\|^2 \leqslant \liminf_n \|\hat{u}_n\|^2 \leqslant \sum_n \operatorname{cap}(A_n) + \varepsilon.$$

另一方面, 由于对所有的 n, $\bar{u} \geqslant u_n$, 在 A_n 的邻域内有 $\bar{u} \geqslant 1$, 从而在 $\bigcup_n A_n$ 的邻域内有 $\bar{u} \geqslant 1$. 所以, $\operatorname{cap}(\bigcup_n A_n) \leqslant \|\bar{u}\|^2$. 结论得证. □

引理 6.49 *设 $f : \Omega \to \mathbb{R}$ 是处处定义的拟连续的可测函数. 若 $f \geqslant 0$ 是几乎处处成立的, 则 $\operatorname{cap}(\{f < 0\}) = 0$.*

证明 令 Ω_n 是拟连续性定义中所涉及的 Ω 中的开子集序列. 因为 f 在 $\Omega \backslash \Omega_n$ 上是连续的, 集合 $\{f < 0\} \backslash \Omega_n$ 是 $\Omega \backslash \Omega_n$ 中的开子集, 因此 $\{f < 0\} \cup \Omega_n$ 是 Ω 中的开集. 令 $f_n \in H_0^1(\Omega)$ 满足 $f_n \geqslant 1$ 于 Ω_n 上几乎处处成立, 则由 $\operatorname{meas}(\{f < 0\}) = 0$, 有 $f_n \geqslant 0$ 于 $\{f < 0\} \cup \Omega_n$ 上几乎处处成立, 即于 $\{f < 0\}$ 的邻域上成立. 因此 $\operatorname{cap}(\{f < 0\} \cup \Omega_n) \leqslant \operatorname{cap}(\Omega_n)$, 从而 $\operatorname{cap}(\{f < 0\}) = 0$. □

由上述引理得到, 若两个拟连续的函数是几乎处处相等的, 则实际上除了在一零容量的集合上外, 它们是相等的. 尤其, 若几乎处处定义的函数是拟连续的, 则在忽略零容量的意义下, 它的拟连续的代表元是唯一的.

引理 6.50 *令 $f \in H_0^1(\Omega)$, 则 (i) f 有拟连续的代表元素, (ii) 存在于 $H_0^1(\Omega)$ 中收敛到 f 的序列 $f_n \in \mathcal{D}(\Omega)$, 且满足对每一 $\varepsilon > 0$, 存在容量小于 ε 的集合 Ω_ε, 使得 f_n 在 $\Omega \backslash \Omega_\varepsilon$ 上一致收敛.*

证明 因为 (i) 可由 (ii) 得到, 只需要证明 (ii). 由于集合 $\mathcal{D}(\Omega)$ 可稠密地嵌入到 $H_0^1(\Omega)$ 中, 存在序列 $f_n \in \mathcal{D}(\Omega)$ 满足 $\|f_n - f\| \leqslant 2^{-n}n^{-1}$, 结果有

$$\sum_n 4^n \|f_{n+1} - f_n\|^2 \leqslant \sum_n 4^n (\|f_{n+1} - f\| + \|f - f_n\|)^2 < +\infty. \tag{6.81}$$

置

$$\Omega'_n := \{\omega \in \Omega : |f_{n+1}(\omega) - f_n(\omega)| > 2^{-n}\};\quad \Omega_n := \bigcup_n^{\infty} \Omega'_k.$$

由容量的定义, $\mathrm{cap}(\Omega'_n) \leqslant 4^n \|f_{n+1} - f_n\|_{1,2}^2$. 引理 6.48 可推出, 当 $n \to \infty$ 时, $\mathrm{cap}(\Omega_n) \to 0$. 另一方面, 若 $\omega \notin \Omega_n$, 则对所有 $k \geqslant n$, 有 $|f_{k+1}(\omega) - f_k(\omega)| \leqslant 2^{-k}$. 于是 f_n 具有在 $\Omega \backslash \Omega_n$ 上连续的一致极限. 结论得证. □

定义 6.51　*称一个性质是拟处处 (q.e.) 成立的, 若它在除了零容量集合外是真的. 例如, $f_n \to f$ 在 Ω 上逐点拟处处成立, 如果除了一零容量集合的所有的 $\omega \in \Omega$, 有 $f_n(\omega) \to f(\omega)$.*

引理 6.52　*令在 $H_0^1(\omega)$ 中 $f_n \to 0$, 用 $\tilde{f}_n$ 记 f_n 的拟连续的代表元, 则存在 $\tilde{f}_n$ 的子列, 拟处处收敛到 0.*

证明　置 $\alpha > 0$. 对每一 n, 根据上述引理, 存在 $\mathcal{D}(\Omega)$ 中的序列 $f_{n,k}$, 它既在 $\mathcal{D}(\Omega)$ 中逐点拟处处收敛于 f_n, 且在 $H_0^1(\Omega)$ 中收敛到 f_n. 因此, 对给定的 $\varepsilon > 0$, 存在 $k(n)$, 满足在 $H_0^1(\Omega)$ 中有 $f_{n,k(n)} \to 0$, 且使得集合

$$A_n := \{\omega \in \Omega : |f_{n,k(n)}(\omega) - \tilde{f}_n(\omega)| > \alpha\}$$

具有小于 $\dfrac{1}{2}\varepsilon$ 的容量. 置 $B_n := \{\omega \in \Omega : |f_{n,k(n)}(\omega)| > \alpha\}$. 因为 $f_{n,k(n)}$ 属于 $\mathcal{D}(\Omega)$, B_n 是开集; 因此在 $H_0^1(\Omega)$ 的意义下, 于 B_n 上有 $f_{n,k(n)} \geqslant \alpha$. 从而有 $\mathrm{cap}(B_n) \leqslant \alpha^{-2}\|f_{n,k(n)}\|_{H_0^1(\Omega)}$. 因此对充分大的 n, $\mathrm{cap}(B_n) \leqslant \dfrac{1}{2}\varepsilon$. 最后, 根据引理 6.48,

$$\mathrm{cap}\{\omega \in \Omega : |\tilde{f}_n(\omega)| > 2\alpha\} \leqslant \mathrm{cap}(A_n \cup B_n) \leqslant \mathrm{cap}(A_n) + \mathrm{cap}(B_n) \leqslant \varepsilon.$$

结论得证. □

现在讨论 $H^{-1}(\Omega)$ 的正锥中的元素作为 Radon 测度的表示, 连续函数的支撑是函数值非零的集合的闭包. 令 $C_{00}(\Omega)$ 记支撑是 Ω 中的紧致集合的连续函数集合, 赋予 $C_{00}(\Omega)$ 类似于 $\mathcal{D}(\Omega)$ 拓扑的引导拓扑, 即给定包含在 Ω 中的紧致集合K, 令 $C_K(\Omega)$ 为支撑在 K 中定义于 Ω 上的连续函数集合. $C_K(\Omega)$ 中的一致收敛的拓扑相联系着桶集类:

$$\mathcal{O}_K := \{g \in C_K(\Omega) : \|g\|_\infty \leqslant 1\},$$

且 $C_{00}(\Omega)$ 是集合类 $C_K(\Omega)$ 的引导极限, 则 $y_n \to y$ 于 $C_{00}(\Omega)$ 中, 如果存在紧致子集 $K \subset \Omega$[①] 满足对所有的 n, y 与 y_n 具有在 K 中的支撑, $y_n \to y$ 是一致的. 赋予这一拓扑, $C_{00}(\Omega)$ 是局部凸的可分的拓扑向量空间. 它在一致收敛拓扑下的闭包是在 $\partial\Omega$ 上消失的在 $\bar{\Omega}$ 上连续的函数空间 $C_0(\Omega)$.

称 L 是在定义于 Ω 上连续函数的给定向量空间 X 上的非负线性形式, 若 L 是 $X \to \mathbb{R}$ 函数, 满足当 $f \in X$[②]且 $f \geqslant 0$, 则有 $L(f) \geqslant 0$.

① 原著中是 $K \in \Omega$.

② 原著中为 $\in X$, 没有 f.

引理 6.53 下述结论成立:

(i) 任何 $C_{00}(\Omega)$ 上的非负线性形式是连续的.

(ii) $H_0^1(\Omega)\cap C_{00}(\Omega)$ 上的非负线性形式在 $C_{00}(\Omega)$ 上具有唯一的非负连续扩展.

证明 (i) 令 L 是 $C_{00}(\Omega)$ 上的非负线性形式. 需要证明集合 $\mathcal{O} := \{y \in C_{00}(\Omega) : |L(y)| \leqslant 1\}$ 是 0 的邻域, 即由引导拓扑的定义, 需要证明, 对 Ω 中的任何紧致子集 K, 对于 $C_K(\Omega)$ 的拓扑, $\mathcal{O}\cap C_K(\Omega)$ 是 0 的邻域. 存在 $y_K \in C_{00}(\Omega)$ 满足于 K 上有 $y_K \geqslant 1$, 处处有 $0 \leqslant y_K \leqslant 1$. 若 $y \in C_K(\Omega)$, 则

$$|L(y)| = |L(y_+) + L(y_-)| \leqslant |L(|y|)| \leqslant \max(y)L(y_K).$$

所以, $\mathcal{O}\cap C_K(\Omega)$ 包含 $C_K(\Omega)$ 的半径为 $L(y_K)^{-1}$ 的球. 证得 (i).

(ii) 令 L 是 $H_0^1(\Omega)\cap C_{00}(\Omega)$ 上的非负线性形式. 如 (i) 的证明, 容易构造作为 $H_0^1(\Omega)\cap C_{00}(\Omega)$ 的元素的函数 y_K. 由同样的论证, 有集合 $\mathcal{O} := \{y \in H_0^1(\Omega)\cap C_{00}(\Omega) : |L(y)| \leqslant 1\}$ 满足 $\mathcal{O}\cap C_K(\Omega)$ 包含 $C_K(\Omega)$ 的半径为 $L(y_K)^{-1}$ 的球与 $H_0^1(\Omega)$ 的交. 由于 $H_0^1(\Omega)\cap C_K(\Omega)$ 是 Banach 空间 $C_K(\Omega)$ 的稠密子集, 根据引理 6.5, 它在 $C_K(\Omega)$ 上有连续的扩展, 依然用 L 表示. 若 $y \in C_{00}(\Omega)$, 则 $L(y)$ 不依赖于包含 y 的支撑的 Ω 的紧致子集 K 的选取. 所以, 这定义了 L 在 $C_{00}(\Omega)$ 上的扩展, 由于 $H_0^1(\Omega)\cap C_{00}(\Omega)$ 是 $C_{00}(\Omega)$ 的稠密子集, 这一扩展是非负的, 证得结论. □

Ω 上的 Borel 测度是 Borelian sigma 代数 $\mathcal{B}_\Omega$ 上的测度. Borel 测度 μ 称为局部有限的, 若只要 K 是 Ω 的紧致子集, 就有 $\mu(K) < \infty$. 我们引用泛函分析中的下述经典结果:

定理 6.54 (Radon-Riesz) 由 $L_\mu(f) := \int_\Omega f(\omega)\mathrm{d}\mu(\omega)$ 定义的映 μ 到泛函 L_μ 的映射是局部有限 Borel 测度集合与 $C_{00}(\Omega)$ 上的非负线性形式集合之间的一对一映射.

令 L 是 $C_{00}(\Omega)$ 上的非负线性形式, 根据上述定理, 与 L 联系着局部有限的 Borel 测度 μ 满足 $L_\mu(f) := \int_\Omega f(\omega)\mathrm{d}\mu(\omega)$. 称与 μ 相联系的 Radon 测度是通过把赋予测度 μ 的 Borel sigma 代数完备化得到的 (这是包含 μ 的 Borel sigma 代数与零测度集合的最小的 sigma 代数.) 用 $M_+(\Omega)$ 记 Radon 测度的集合. 每一 Radon 测度是正则的.

令 $\mu \in H^{-1}(\Omega)_+$, 则 μ 在 $H_0^1(\Omega)\cap C_{00}(\Omega)$ 上的限制是非负线性形式, 根据引理 6.53, 它可以扩展为 $C_{00}(\Omega)$ 上的非负线性形式. 根据 Radon-Riesz 定理6.54, 这一扩展是局部有限的 Radon 测度, 仍然用 μ 记它. 所以, 可以将 $H^{-1}(\Omega)_+$ 与 $H^{-1}(\Omega)\cap M_+(\Omega)$ 等同起来.

引理 6.55 令 A 是 Borel 集合, 则 A 具有零容量的充分必要条件是对所有的 $\mu \in H^{-1}(\Omega)\cap M_+(\Omega)$, 有 $\mu(A) = 0$.

证明 若 A 具有零容量, 则对给定的 $\varepsilon>0$, 存在 $u\in H_0^1(\Omega)$ 满足 $\|u\|_{H_0^1(\Omega)}\leqslant\varepsilon$, 且以 $H_0^1(\Omega)$ 的意义于 A 上有 $u\geqslant 1$, 即存在开集 $A_\varepsilon\supset A$ 满足 $u\geqslant 1$, 于 A_ε 上是几乎处处成立的, 则存在连续函数 f, 在 $[0,1]$ 上取值, 其支撑在 A_ε 中, 满足 $\mu(A_\varepsilon)\leqslant\mu(f)+\varepsilon$, 可设 $f\in H_0^1(\Omega)$. 因为 $u-f\in H_0^1(\Omega)_+$, 有

$$\mu(A)\leqslant\mu(A_\varepsilon)\leqslant\langle\mu,f\rangle_{H^{-1}(\Omega),H_0^1(\Omega)}+\varepsilon$$
$$\leqslant\langle\mu,u\rangle_{H^{-1}(\Omega),H_0^1(\Omega)}+\varepsilon\leqslant\varepsilon(1+\|\mu\|_{-1.2}).$$

因为 ε 可取任意小, μ 是非负的, 可得 $\mu(A)=0$.

相反地, 设 A 具有一定正的容量. 令

$$T:=\{u\in H_0^1(\Omega): \text{以在 } H_0^1(\Omega) \text{ 的意义于 } A \text{ 上 } \mu\geqslant 1\}.$$

则对某 $\varepsilon>0$, 集合 T 与 $S:=\{u\in H_0^1(\Omega):\|u\|<\varepsilon\}$ 是不相交的. 根据分离定理 2.13, 存在非零的 $\mu\in H^{-1}(\Omega)$ 满足

$$\langle\mu,u'\rangle\leqslant\langle\mu,u\rangle,\quad\forall u'\in S,\quad u\in T.$$

对 $u'\in S$ 极大化左端, 得到

$$\varepsilon\|\mu\|_{-1,2}\leqslant\langle\mu,u\rangle,\quad\forall u\in T.\tag{6.82}$$

要证 μ 是非负的. 令 $u\in H_0^1(\Omega)_+$, $u_0\in T$, 则对所有的 $\alpha>0$, 有 $\alpha u+u_0\in T$, 因而

$$\langle\mu,u\rangle=\lim_{\alpha\uparrow\infty}\alpha^{-1}\langle\mu,\alpha u+u_0\rangle\geqslant 0.$$

根据引理 6.53(ii), 在具有唯一的非负连续扩展的意义下, u 是 $H^{-1}(\Omega)\cap M_+(\Omega)$ 的元素, 则由 (6.82), 因为 μ 是正则的 Borel 测度, 有 $\mu(A)\geqslant\varepsilon\|\mu\|_{-1,2}>0$. 结论得证. □

引理 6.56 令 $f\in H_0^1(\Omega)$, 具有拟连续的代表元 $\tilde f$. 如果 $\mu\in H^{-1}(\Omega)\cap M_+(\Omega)$, 则 $\tilde f\in L_1(\mu)$ 且

$$\int_\Omega\tilde f(\omega)\mathrm{d}\mu(\omega)=\langle\mu,f\rangle_{H^{-1}(\Omega),H_0^1(\Omega)}.$$

证明 根据引理 6.50, 存在 $f_n\in\mathcal{D}(\Omega)$ 满足于 $H_0^1(\Omega)$, 且在 Ω 上为拟处处均有 $f_n\to\tilde f$, 则 f_n 在 $H_0^1(\Omega)$ 与 $L_1(\mu)$ 中均是 Cauchy 序列. 根据引理 6.55, f_n 的拟处处收敛性可推出 μ 几乎处处收敛性, 可得 $\tilde f$ 对于 μ 是可测的, 它是 f_n 在 $H_0^1(\Omega)$ 与 $L_1(\mu)$ 中的极限. 由于要验证的等式对每一 f_n 均是成立的, 结论得证. □

最后给出 $H_0^1(\Omega)$ 中的 $H_0^1(\Omega)_+$ 的切锥与法锥的公式. 置 $M_-(\Omega):=-M_+(\Omega)$.

定理 6.57 令 $f \in H_0^1(\Omega)_+$, 有拟代表元素 $\tilde{f}$, 则 $H_0^1(\Omega)_+$ 在 $f \in H_0^1(\Omega)_+$ 处的切锥与法锥是

$$T_K(f) = \{g \in H_0^1(\Omega)_+; g \geqslant 0 : \text{q. e. 在 } \{\tilde{f} = 0\} \text{上}\}, \tag{6.83}$$

$$N_K(f) = \{\mu \in H^{-1}(\Omega) \cap M_-(\Omega) : \ \mu(\{\tilde{f} > 0\}) = 0\}. \tag{6.84}$$

证明 置 $K := H_0^1(\Omega)_+$. 因为 K 是凸锥, 由 (2.110), 有 $N_K(f) = K^{-1} \cap (f)^{\perp}$. 我们知道 $K^- = -K^+ = H^{-1}(\Omega) \cap M_-(\Omega)$. 令 $\mu \in K^-$. 由引理 6.49, $f \geqslant 0$ 是拟处处成立的, 根据引理 6.55, 有 $\mu(\{\tilde{f} < 0\}) = 0$. 因为由引理 6.56, $\tilde{f} \in L_1(\mu)$, 且 $-\mu \in M_+(\Omega)$, 根据 6.56, 有 $\langle \mu, f \rangle = 0$ 当且仅当 $\mu(\{\tilde{f} > 0\}) = 0$. 于是有 N_K 的公式成立.

剩下要证 (6.83). 令 $\mu \in N_K(f)$ 与 $g \in H_0^1(\Omega)$. 由引理 6.49 与 6.55, $\mu(\{\tilde{f} < 0\}) = 0$, 则由 (6.84) 与 $\tilde{f}$ 的 μ 可测性, 有

$$\mu(g) = \int_{\{\tilde{f}=0\}} g(\omega) \mathrm{d}\mu(\omega). \tag{6.85}$$

因为拟连续函数 $\tilde{g}$ 本身也是 μ 可测的, 得到

$$\mu(g) = \int_{\{\tilde{f}=0\} \cap \{\tilde{g}<0\}} g(\omega) \mathrm{d}\mu(\omega) + \int_{\{\tilde{f}=0\} \cap \{\tilde{g} \geqslant 0\}} g(\omega) \mathrm{d}\mu(\omega). \tag{6.86}$$

若 g 属于 (6.83) 的右端的集合, 由引理 6.55, 上式的第一项积分是零. 与 (6.84) 相结合, 可得 $\mu(g) \leqslant 0$, 因此 g 属于 $T_K(f)$.

相反地, 设集合 $T := \{\tilde{f} = 0\} \cap \{\tilde{g} < 0\}$ 具有非零的容量. 由引理 6.48, 存在 $\varepsilon > 0$ 满足集合 $T_\varepsilon := \{\tilde{f} = 0\} \cap \{\tilde{g} < -\varepsilon\}$ 具有非零的容量. 由引理 6.55, 存在 $\mu \in H^{-1}(\Omega) \cap M_+(\Omega)$ 满足 $\mu(T_\varepsilon) > 0$. 则由 (6.84) 可得, 由

$$\tilde{\mu}(g') := -\int_{T_\varepsilon} g'(\omega) \mathrm{d}\mu(\omega), \quad \forall g' \in H_0^1(\Omega) \tag{6.87}$$

定义的 $\tilde{\mu} \in H^{-1}(\Omega)$ 是满足 $\tilde{\mu}(g) \geqslant \varepsilon_\mu(T_\varepsilon) > 0$ 的 $N_K(f)$ 的元素. 所以, g 不属于 $T_K(f)$. □

6.4.4 灵敏度分析与最优控制

由于 $H_0^1(\Omega)$ 的正锥是多面性的, 为了得到障碍问题的最优解关于控制变量的方向导数, 可以应用 4.7.2 节的结果.

在后面, 用 $y(u)$ 记障碍问题解的拟连续表示元素. 记 $\lambda(u) := u + \Delta y(u)$ 为与约束 $y(u) \in H_0^1(\Omega)_+$ 相联系的乘子. 根据定理6.44, 由于 $\lambda(u) \in L_2(\Omega)$, 用定理 6.57, 得

到临界锥的下述表达式

$$C(u) := \left\{ z \in H_0^1(\Omega):\ z \geqslant 0 \text{ q. e. } 在 \{y(u)=0\}上;\ \int_\Omega \lambda(u)(\omega)z(\omega)\mathrm{d}\omega = 0 \right\},$$

或等价地,

$$C(u) := \{ z \in H_0^1(\Omega):\ z \geqslant 0 \text{ q. e. } 在 \{y(u)=0\}上;\ z(u+\Delta y(u)) = 0 \text{ a.e.}\}.$$

子问题的表达式是

$$\min_{z\in C(u)} \int_\Omega |\nabla z(\omega)|^2 \mathrm{d}\omega - 2\int_\Omega v(\omega)z(\omega)\mathrm{d}\omega.$$

这一具有强凸目标函数的凸问题具有唯一的解, 可由相应的最优性系统刻画

$$\int_\Omega \nabla z(\omega)\cdot\nabla z'(\omega)\mathrm{d}\omega \geqslant \int_\Omega v(\omega)z'(\omega)\mathrm{d}\omega, \quad \forall z' \in C(u),\ z \in C(u). \tag{6.88}$$

注意到 $u \to y_u$ 为 $L_2(\Omega) \to L_2(\Omega)$ 的 Lipschitz 连续函数, 由它的方向可微可推出 Hadamard 方向可微, 下述结果是定理 4.91 的一个简单结论.

定理 6.58　*从 $L_2(\Omega)$ 到 $H_0^1(\Omega)$ 的映射 $u \mapsto y(u)$ 是 Hadamard 方向可微的, 沿方向 v 的方向导数 z 是 (6.88) 的唯一解.*

对给定的 u, 下面讨论 $y(u)$ 的 Hadamard 方向可微性的充分条件. 注意到, 这里引入的严格互补性的概念并不是定义 3.13 中引入的抽象概念的特殊情况.

定义 6.59　*称 $y(u)$ 满足严格互补性条件, 如果*

$$C(u) = \{z \in H_0^1(\Omega) : z = 0 \text{ q.e. } 在 \{y(u)=0\}上\}.$$

在这一假设下, 临界锥是向量空间. 所以, (6.88) 等价于

$$\int_\Omega \nabla z(\omega)\cdot\nabla z'(\omega)\mathrm{d}\omega = \int_\Omega v(\omega)z'(\omega)\mathrm{d}\omega, \quad \forall z' \in C(u), z \in C(u). \tag{6.89}$$

推论 6.60　*在定理 6.58 的假设之下, 若对某一 u, 严格互补性条件成立, 则映射 $y(\cdot) : L_2(\Omega) \to H_0^1(\Omega)$ 是 Hadamard 可微的, 沿方向 v 的方向导数是 (6.89) 的唯一解.*

前述结果使我们能够计算状态与控制的函数, 如由 (6.18) 定义的二次目标函数 $J_\tau(y,u)$ 的方向导数. 像以前一样, $N>0$ 与 $\tau \in L_2(\Omega)$ 是给定的.

定理 6.61　(i) *映射 $F(u) := J_\tau(y(u),u)$, 其中 $y(u)$ 记变分不等式(6.77) 的解, 在 u 处是 Hadamard 方向可微的, 沿方向 v 的方向导数为*

$$F'(u;v) = \int_\Omega (y(u)(\omega) - \tau(\omega))z(\omega)\mathrm{d}\omega + N\int_\Omega u(\omega)v(\omega)\mathrm{d}\omega, \tag{6.90}$$

其中 z 是 $y(u)$ 沿方向 v 的方向导数[①].

(ii) 若还有 u 满足严格互补条件, 则 $F(\cdot)$ 是于 u 处为 Hadamard 可微的, 其导数是 $p_u + Nu$, 其中 $p \in H_0^1(\Omega)$ 是 $C(u)$ 中下述方程的唯一解

$$\int_\Omega \nabla p_u(\omega) \cdot \nabla z'(\omega) \mathrm{d}\omega = \int_\Omega (y_u(\omega) - \tau(\omega)) z'(\omega) \mathrm{d}\omega, \quad \forall z' \in C(u). \tag{6.91}$$

证明 结论 (i) 由定理 6.58 与命题 2.47 的运算法则可得到. 若严格互补条件成立, 先观察到 (6.91) 具有唯一解, 因为它是具有强凸目标函数的最优化问题

$$\min_{p \in C(u)} \frac{1}{2} \int_\Omega |\nabla p(\omega)|^2 \mathrm{d}\omega - \int_\Omega (y_u(\omega) - \tau(\omega)) p(\omega) \mathrm{d}\omega$$

的最优性系统. 若 u 满足严格互补条件, 则由推论 6.60 可得, z 是 (6.89) 的解. 于是

$$\int_\Omega (y_u(\omega) - \tau(\omega)) z(\omega) \mathrm{d}\omega = \int_\Omega \nabla p_u(\omega) \cdot \nabla z(\omega) \mathrm{d}\omega = \int_\Omega p_u(\omega) v(\omega) \mathrm{d}\omega,$$

从而可得到 (ii). □

现在考虑最优控制问题

$$\text{(P)} \qquad \min_u J_\tau(y(u), u), \quad \text{s.t.} \quad u \in K, \tag{6.92}$$

其中 K 是 $L_2(\Omega)$ 中的闭凸子集合.

定理 6.62 令 u 是上述最优控制问题的局部解, 则下述结论成立:

(i) 对所有的 $v \in K$, $F'(u, v-u) \geqslant 0$.

(ii) 若还有 u 满足严格互补条件, 则

$$\int_\Omega (p_u(\omega) + Nu(\omega))(v(\omega) - u(\omega)) \mathrm{d}\omega \geqslant 0, \quad \forall\, v \in K.$$

证明 由于 u 是局部极小点, 对所有的 $t \in [0,1]$, $u + t(v-u) \in K$, 有

$$0 \leqslant \lim_{t \downarrow 0} t^{-1}(F(u + t(v-u)) - F(u)) = F'(u, v-u).$$

结论 (i) 得证. 将此结论与定理 6.62(ii) 相结合, 证得结论 (ii). □

注 6.63 这一节的结果可推广到 Laplace 算子的情况, 考虑一般的二阶椭圆算子, 即当 $y(u)$ 由下述变分不等式的解给出的情况

$$\sum_{i,j=1}^n \int_\Omega a_{ij}(\omega) \frac{\partial y}{\partial \omega_i}(\omega) \cdot \frac{\partial y'}{\partial \omega_i}(\omega) \mathrm{d}\omega \geqslant \int_\Omega v(\omega) y'(\omega) \mathrm{d}\omega, \quad \forall y' \in H_0^1(\Omega),\ y \in H_0^1(\Omega). \tag{6.93}$$

① 原著中上式等号左端为 $F(u, v)$.

设 $a(\cdot)$ 在 Ω 的闭包上是二阶连续可微的, $a(\cdot)$ 的椭圆性意味着: 对某一 $\alpha > 0$, 所有的 $\omega \in \Omega$ 与 $x \in \mathbb{R}^n$, 成立

$$\sum_{i,j=1}^{n} a_{i,j}(\omega) x_i x_j \geqslant \alpha \sum_{i=1}^{n} x_i^2, \tag{6.94}$$

可以证明, 对每一 $u \in L_2(\Omega)$, (6.93) 在 $H^2(\Omega) \cap H_0^1(\Omega)$ 中有唯一的解 $y(u)$. 可以结合关于 $H_0^1(\Omega)_+$ 的多面性的上述结果与 5.1.1 节中抽象变分不等式的分析, 对 $u \mapsto y(u)$ 的灵敏度进行分析. 尤其, 结合定理5.5, 可得到映射 $u \mapsto y(u) : L_2(\Omega) \to H_0^1(\Omega)$ 是 Lipschitz 连续的且 Hadamard 方向可微的, $y(u)$ 沿方向 v 的导数 z 是下述切化变分不等式的唯一解

$$\sum_{i,j=1}^{n} \int_{\Omega} a_{ij}(\omega) \frac{\partial z}{\partial \omega_i}(\omega) \cdot \frac{\partial z'}{\partial \omega_i}(\omega) \mathrm{d}\omega \geqslant \int_{\Omega} v(\omega) z'(\omega) \mathrm{d}\omega, \quad \forall z' \in C(u), \ z \in C(u). \tag{6.95}$$

第7章 文献注记

7.1 背景素材

关于基本泛函分析的素材是经典的，至少对 2.1.1 节与 2.1.3 节是这样的，素材的详细论述可在各个教科书中找到[67,213]. 基于 Zorn 引理的 Hahn-Banach 定理(定理 2.10) 的证明，可以参看文献 [213].

Banach 空间或更一般的局部凸的拓扑向量空间中方向可微性各种概念的详尽讨论可参考文献 [16,159,191]. 增广实值函数的上图导数的一个全面处理可参看文献 [9,184].

Clarke 切锥的概念由 Clarke 提出，并在文献 [57] 中给出了详细讨论. Clarke 切锥等于余切锥的下上图极限 (方程 (2.92)) 这一事实，由 Penot[163] 证得; 这里的证明取自文献 [184, 定理 6.26]. 闭集合 S 称为在点 $x \in S$ 处是 Clarke 正则的，若 S 在 x 处的 Clarke 切锥与余切锥彼此相等.

广义开映射定理(定理 2.70) 来自 Robinson[173] 与 Ursescu[207]; 我们采用 Robinson[173] 的证明. 引理 2.74 来自于 Robinson[173]; 我们采用文献 [38] 的证明.

2.3.2 节的素材中的大部分取自于文献 [99]. 表明开性与度量正则性概念的等价性的定理2.81 取自于文献 [99]，更早的证明在文献 [103] 中出现. 度量正则性的概念已经在 Robinson[173] 的工作中被显式地给出了，其他几位作者也发展过这一概念. 它的源头可追踪到 Lyusternik[143] 与 Graves[88] 的经典工作与 Hoffman 引理的结果. 关于这一概念进展的一个历史的评述可在 Ioffe 的最近一篇文章[100] 中找到.

稳定性定理 2.83 的“充分性”部分是经典的 Robinson[173]-Ursescu[207] 稳定性定理. 给出扰动的多值函数是度量正则的充分条件的定理 2.84 是非线性分析中长期发展的结果. 其证明可追溯到 Banach 不动点定理2.2 与 Lyusternik[143] 及 Graves[88] 的工作. 度量正则性比率估计 $c(1-c\kappa)^{-1}$ 出现在文献[102]; 还可参考文献 [60, 定理 2.1]. 正则性条件 (2.163) 与稳定性定理 2.87 来源于文献 [174]. 命题 2.90 源于文献 [174, 推论 2]. Mangasarian-Fromovitz 约束规范在文献 [149] 中给出叙述. 集值分析的广泛的研究与讨论由文献 [14] 给出.

凸函数理论的完整讨论由 Rockafellar[179] (有限维情况), Ekeland and Teman[70], Aubin and Ekeland[13] 与 Ioffe and Tihomirov[106] 给出. 近似次微分的理论可用来对最优解集为空集的凸优化问题的最优值函数进行灵敏度分析[95]. 关于二阶近似次微分的内容可参看文献 [94,96].

对偶性理论在最优性理论中是经典的. Lagrange 对偶性的源头在 von Neumann 的对策论中, 之后被许多作者发展. 共轭对偶性(2.5 节中给出) 主要来自文献 [180]. 命题 2.141 是由 Cominetti 指出的 (私人讨论). 定理 2.144 源于文献 [180]. 命题 2.152 来源于文献 [173, 推论 1]. 定理 2.182 的证明取自于文献 [179]. 导致平静性概念的最优值函数次可微性质与对偶问题性质间的联系在文献 [179,180] 中显式地给出. 平静性 (calmness) 这一术语由 Clarke[56] 建议使用. 实际上, 在文献 [56] 中, 非扰动问题(P) 被称为是平静的, 如果对充分大的 $r>0$, 函数 $v(u)+r\|u\|$ 在 $u=0$ 处取到极小值. 根据命题 2.148, 在凸的情形, 平静性的定义等价于定义 2.146 所给出的.

注 2.172 推广了定理 2.171, 来源于 Attouch and Brézis[10].

锥线性问题的对偶性在文献 [7] 中讨论. 例 2.194 来源于 Kretschmer[123], 还可参看文献 [7,pp.43-45]. 次相容的锥线性问题的概念由文献 [7] 通过序列的形式定义, 如命题 2.188 之后所讨论的那样. 由 2.5.1 节给出的次值lsc $v(u)$ 与一次相容问题(P_u) 的概念是这一概念的一个自然的抽象框架.

当然, 线性规划问题的“强对偶”性质是周知的. 关于广义线性规划问题的处理似乎是新的. Hoffman 引理(定理 2.200) 对于有限维空间而言, 由文献 [97] 导出, 它被推广到无限维情况, 如文献 [102]. 多面体多值函数的定义 2.206 是 Robinson[176] 所引入的这一概念到 Banach 空间的自然推广. 线性规划问题关于右端项扰动的最优解集的 Lipschitz 连续性 (如, 例 2.211) 由 Mangasarian 和 Shiau[150] 讨论.

7.2　最优性条件

函数的导数在其无约束最优解处为零这一基本事实是著名的 Fermat 引理. 变分问题的 (一阶与二阶) 最优条件理论主要由 Euler, Lagrange, Jacobi, Weierstrass 与 Legendre 发展起来. Lagrange 乘子由 Lagrange 在他关于力学的论文中引入 [130]. 非线性规划的一阶最优性条件由 Karush[113] 与 Kuhn and Tucker[125] 表述, 在数学规划的现代文献中经常被称为 KKT 最优性条件.

与非线性规划问题的局部最优解相联系的广义 Lagrange 乘子的存在性由 John[11] 得到. 在 John 的结果很久之前, 在变分计算中, 广义 Lagrange 乘子的存在性结果已经是周知的了. 另一方面, 变分计算中的具体特点遮蔽了用于研究最优性的某些方法的一般性的本质.

形式为 (3.1) 的抽象约束的最优化问题的一阶最优性条件在 20 世纪 70 年代由 Robinson[172], Kurcyusz[126] 与 Zowe and Kurcyusz[215] 研究. 3.11 节采用的对偶性方法受 Rockafellar[179,180] 的工作所启发. 保证广义 Lagrange 乘子集合非空的充分性条件 (类似于命题 3.18 的条件) 由 Zowe 和 Kurcyusz[215] 导出. Lagrange 乘子集合非空且有界的充分必要条件是 Robinson 约束规范成立这一结果 (见命题 3.17),

在非线性规划的情况由 Gauvin[76] 推导, 在集合 K 具有非空内部的假设的情况由 Zowe 和 Kurcyusz[215] 得到. 来自于 Ekeland[69] 的 Ekeland 变分原理 (定理 3.22) 的详细讨论, 可在文献 [13] 中找到.

如果相应的最优化问题有唯一极小点, 在 “sharp 极小” 概念之下, 一阶增长性概念在逼近论中是周知的. 在 “弱 sharp 极小” 的概念下, Burke 和 Ferris[52] 研究了一般性的一阶增长条件. Banach 空间中的一阶 (还有二阶) 充分性条件在文献 [151] 中用近似临界锥与 (非广义的)Lagrange 乘子表述. 例 3.25 来自于 Maurer 和 Zowe[151].

非线性规划的二阶最优性条件的一个形式已经出现在文献 [74] 中. 非线性规划的二阶最优性条件在 20 世纪 70 年代已被前苏联文献所研究[136]. (在 Mangasarian-Fromovitz 约束规范下) 非线性规划问题的无间隙二阶最优性条件由 Ioffe[101] 与 Ben-Tal[23] 给出.

对于半无限规划, (在非常严格的假设下) 二阶最优性条件首先在文献 [93,212] 中用所谓的简约化方法推导. 在这些文章中, 为得到 “无间隙” 二阶条件, 表示 K 的曲率的额外项已经很清楚地出现了. 试图以抽象的形式描述这一额外项 (在半无限规划的情形) 由 Kawasaki[114] 给出. 还有强有力的研究分布于各个文献中, 旨在消除二阶必要性条件与二阶充分性条件间的间隙 [51,60,104,105,161,165,166,183].

用 (抛物) 二阶导数的概念推导二阶最优性条件起源于 Ben-Tal[23] 与 Ben-Tal 和 Zowe[24]. 带有与内二阶切锥相对应的 sigma 项的 Lagrange 形式的二阶必要性条件源于 Cominetti[61]. 由定理 3.45 给出的二阶必要性条件在文献 [40,164] 中叙述, 其中 sigma 项联系着外二阶切集的凸子集. 例 3.35 是新的, 非凸二阶切集的这样的例子以前似乎是没有的. 命题 3.34 源于 Cominetti[60].

在 Hilbert 空间框架下多面性 (定义 3.51) 的概念由 Mignot[152] 和 Haraux[89] 引入, 被用于计算到凸集的投影的方向导数问题. 这里, 在一般 Banach 空间下多面性的处理似乎是新的. 在 Hilbert 空间的框架下, 定理 3.58 的结果出现在文献 [89] 中. 增广多面性条件 (定义 3.52) 源于文献 [38, Part I], 在那里被用于灵敏度分析. 它在最优性条件理论的应用是近期的事情[34]. Legendre 形式 (定义 3.73) 是变分计算中的经典概念. 3.3.2 节的表述取自于文献 [106]. 广义Legendre 形式的概念是新的. 集合的二阶正则性的概念由 Bonnans 和 Cominetti[38] 给出, 由 Bonnans, Cominetti 和 Shapiro[40,39] 系统发展起来. 具有相等的非空的二阶切集但并不是二阶正则的集合的例 3.87 是新的. 二阶上图正则函数的概念 (定义 3.94) 是新的.

二阶次导数在文献 [105,183,184] 中被系统地研究. 命题 3.100 与抛物正则性的概念取自 Rockafellar 和 Wets[184]. 命题 3.103 的结果由 Rockafellar 建议 (私人交流). 用复合最优化方法推导的最优性 (一阶与二阶) 条件被 Ioffe[101,105], Penot[166] 与 Rockafellar[182] 讨论.

精确惩罚函数在非线性规划问题的情况由 Eremin[71] 引入, 尽管对线性规划问题而言, 在大 "M" 方法的名义下, 它们被熟知了很长时间. 具有小的非光滑项的精确惩罚函数在文献 [31] 中讨论, 它们同增广对偶性的关系由 Rockafellar[181] 讨论. 命题 3.111 的结果来源于文献 [101]. 定理 3.115 的源于 Clarke[56], 见文献 [50]. 用于算法程序的设计, 增广 Lagrange 函数由 Hestenes[91] 与 Powell[169] 引入. 非线性规划的精确惩罚函数与增广 Lagrange 函数的一般性处理由 Bertsekas[28] 给出.

线性规划的严格互补的定理 3.133 由 Goldman 和 Tucker[84] 给出. 有限维二次规划解的存在性由 Frank 和 Wolfe[75] 得到. 到无限维情况的推广 (定理 3.128) 是新的. 在凸二次不等式约束与费用函数的情形, 由 Luo 和 Zhang[142] 证得, 有限值问题具有一解. 二阶充分性条件 (定理 3.130) 来源于 Majthay[144], 还可参看文献 [62, 148].

只有有限几篇文献讨论过非孤立极小点的二阶最优性条件. 二阶增长条件的必要性条件是孤立极小点的相应条件显然的推广. Shapiro 的早期文章[189] 讨论了严格互补条件下的非线性规划问题, 将其用于对称矩阵的最大特征值的灵敏性分析. Bonnans 和 Ioffe 在文献 [41] 中建立了凸规划二阶增长条件的刻画, 文献 [42] 在严格的假设下, 用追近法向量对非线性规划的二阶增长条件给出刻画. 这里根据这些文献给出表述. 在非线性规划的情形, Burke 和 Ferris[52], Ward[210,211] 与 Studniarski 和 Ward[205] 给出了二阶增长性的一些充分性条件. 除了二阶正则性的讨论 (定理 3.150) 由 Cominetli 与作者[40] 给出, 到带有一般约束的最优化问题的推广是新的.

7.3　稳定性与灵敏度分析

与研究极大 - 极小问题相联系的最优值函数的 (一阶) 可微性质的研究可追溯到 Chebyshev 关于代数多项式一致近似的经典工作, 也是 20 世纪六、七十年代非滑分析系统进展的动因之一. 有很多工作用于研究极大 - 极小最优化的不同侧面[64,65]. 将经典的隐函数定理用于表示为 (非线性) 方程组形式的一阶最优性条件得到的最优解的可微性性质由文献 [74] 给出. 这一方法较详尽的讨论及相关的参考文献可参看文献 [72]. 基于方程组的方法在文献 [175,177] 的工作中被扩展和推广到用于变分不等式(广义方程) 的灵敏度分析中.

70 年代某些开创性的工作是前苏联取得的, 如文献 [66,133,134] (额外的参考文献还可参看文献 [135]). 遗憾的是只有一些支离破碎的结果被西方所知, 在这一专题上的进展几乎没有或者根本没有对西方世界产生影响.

通过推导相应的方向导数的上方与下方估计研究最优值函数的一阶可微性性质的思想至少可追溯到 Danskin[64]. 最近几年, 结合这一方法与二阶分析, 在最优解的渐近性态方面的研究取得了本质性的进展. 这一方法的主要思想基于最优值函

数二阶的上方与下方估计. 当这些估计彼此充分接近时, 可以推导出最优值函数的二阶展开式, 从而给出最优 (近似最优) 解的展开的公式. 得到的最优解的一阶展开式是通过辅助的最优化问题描述的. 这一思想隐含地出现在文献 [85] 中, 但在文献 [78, 79, 187, 190] 中就已经很清晰了. 另一个重要的进展由 Auslender 和 Cominetti[15] 给出, 它通过用对偶性结果与二阶切集联系起来. 这一章所采用的方法正是后者.

4.1 节给出的连续性结果是周知的[17]. 函数的上图极限与命题 4.6 形式的结果在文献 [11] 中被详细讨论.

4.2 节中定义的方向正则性的概念, 在非线性规划的情况, 在文献 [85] 中引入. 在文献 [38, part I] 被推广到 Banach 空间下的抽象框架下, 其中定理 4.9 与推论 4.10 就是那里导出的.

最优值函数的一阶可微性质的研究得到人们极大的关注. 定理 4.13 本质上来源于 Danskin[64], 那里最优值函数 $v(u)$ 的方向导数公式 (4.33) 是在有限维的情况得到的. 令人吃惊的是, 在定理 4.13 的假设之下, $v(u)$ 是 Frechét 意义下方向可微的. 定理 4.17 源于 Levin[132] 与 Valadier[208] (还可参考文献 [106, 定理 3, pp.201]).

可行集也发生扰动的情况下的最优值函数的一阶分析比可行集固定情况要棘手得多. 例 4.23 表明, 要得到最优值函数的一阶导数, 相应最优化问题的线性化是不足够的, 这一例子源于 Gauvin 和 Tolle[80]. 在 Robinson 约束规范下, 命题 4.22 由 Lempio 和 Maurer[131] 证得, 当前的形式由文献 [38, part I] 给出的. 处理凸情形的定理 4.24, 由结合文献 [38, 命题 3.2, Part I] 与文献 [197] 的技术得到, 它是 Gol'shtein 定理[86] 到 Banach 空间情形下的推广. 定理 4.25 是 Levitin[133,134] 与 Lempio 和 Maurer[131] 给出的结果到方向正则性框架下的推广. 针对非线性规划问题的定理 4.26 由 Gauvin 和 Tolle[80] 叙述 (也可参看文献 [77]), 在文献 [131] 中被推广到 Banach 空间的情况.

最优解与 Lagrange 乘子的量化稳定性的不同侧面被许多学者研究. 非光滑情形下, 凸优化问题解的 Lipschitz 性质被 Aubin[12] 与 Janin 和 Gauvin[109] 研究. 当可行集固定时, 导致其最优解的 Lipschitz 估计的由命题 4.32 中给出的简单上界源于 Shapiro[192]. 在不同的一般性的框架下, 最优解的 Hölder 稳定性被许多文献给出[5,6,43,38,78,192]. 命题 4.37 与 4.41 取自文献 [192].

严格约束规范的概念 (定义 4.46) 由 Shapiro[192] 引入, 命题 4.47 的相应结果也在该文中证明. 定理 4.51 取自文献 [196]. 在集合 K 为多面体的情形, 定理 4.51 的 Lipschitz 稳定性结果源于 Robinson[177]. 例 4.54 取自于 Shapiro[198]. 在非线性规划的情形, 命题 4.52 的结果源于 Anitescu[8].

定理 4.55 中的方向 Lipschitz 稳定性结果是理论发展过程中的一个结果. 在非线性规划的情况, 在二阶充分性条件的强形式下, 最优解的 Lipschitz 稳定性由 Robinson[177] 通过广义方程方法建立. 在非线性规划的情况下, 二阶充分性条件

(4.138) 由 Shapiro[190] 引入, 在该文献中给出论断, 这样的二阶条件可能是方向 Lipschitz 稳定性所需要的最弱的条件. 在方向正则性条件下, 这一结果在文献 [15] 中被进一步推广到文献 [201] 中的 Banach 空间的情况. 集合约束问题的一般情形所需要的额外条件 (4.142) 在文献 [195] 中引入. 这里给出的最后形式取自于 Bonnans, Cominetti 和 Shapiro[39]. 由引理 4.57 给出的最优值函数的上方估计源于文献 [195]. 定理 4.60 来自于文献 [43,38]. 例 4.63 受 Janin 和 Gauvin[109] 的一个例子的启发得到的.

4.6 节中的素材, 其中的稳定性分析通过简约方法得到, 基于文献 [44]. 横截性概念的讨论可参考文献 [87]. 对于多面体集合, 非退化性的条件 (4.180) 与 Robinson[178] 用到的是重合的. 在非线性规划的情形, 即当集合 K 是多面体时, 定理 4.81 的稳定性结果由 Shapiro[189] 推导 (在相对严格的假设下), Klatte[119] 通过不同的技巧得到.

4.7.1 节中的素材, 其中最优值函数的上二阶估计中含有 sigma 项, 基于文献 [38]. 没有 sigma 项的相应的下方估计在文献 [15,30,38,78,190] 中不同程度的被推导出来. 在非线性规划的情况, 由定理 4.94 与定理 4.95 给出函数的二阶方向导数与最优解的一阶方向导数, 在 MF 约束规范下由 Shapiro[190] 得到, 在方向正则性条件之下由 Auslender 和 Cominetti[15] 得到. 例 4.99 取自于 Shapiro[190], 它基于 Robinson[177] 的一例子.

4.7.3 节的素材, 其中下二阶估计与相应的二阶展开包含有 sigma 项, 在二阶正则条件下可消除间隙, 基于 Cominetti 与作者的工作[39].

Hölder 稳定情形的扰动性分析源于 Gollan[85] 的工作, 给出了非光滑非线性规划的最优值函数变化的一阶估计. 最优值函数的导数的一个公式由 Gauvin 和 Janin[78] 给出. 近似解展开的一般性公式来自于 Ioffe 与作者[43]. 到 Banach 空间情形的推广由 Bonnans 和 Cominetti[38] 实现, 他们得到了一个一般性的上方估计, 研究了广义多面性的情况, Cominetti 与作者[39] 研究了二阶正则的情形. 然而, 这些文献都假设严格的条件, 被称为 "强方向规范条件", 而这里我们仅假设方向正则性条件成立.

当可行集是固定的, 4.9.2 节中关于最优值函数的一致二阶近似的结果是新的. 定理 4.136 取自于文献 [192]. 4.93 节基于文献 [189].

二阶切集的计算被 Cominetti 和 Penot[61] 讨论, 因此, 4.10.1 与 4.10.2 节中大部分素材基于文献 [61]. 定理 4.149 源于 Cominetti 和 Penot[61] . 命题 4.151 和 4.10.3 节中的素材是新的.

灵敏度分析的不同的处理方法

本书试图给出关于最优化问题的扰动理论的具有一致性的处理. 然而, 值得指出的是, 这里没有讨论最优化问题的灵敏性分析的不同处理方式的最新进展. 一个

这样的方式基于非凸法锥、次微分与协导数 (coderivative) 的概念, 这涉及对偶空间的广义微分的构造. 这一方式中用到的次微分 (法锥) 协导数构造是非凸值的, 因此不与任何切生成的构造是对偶的. 这些非凸构造的理论由 Mordukhovich[154,155] 发展起来.

可以用非凸的协导数得到一般非凸多值函数[156] 的开性、度量正则性与 Lipschitz 稳定性的等价性质的必要与充分性条件. 这些结果的各种无穷维推广由 Mordukhovich 和 Shao[158], Jouran 和 Thibault[112] 与 Ioffe[99] 得到. 到最优化问题与广义方程(变分不等式) 的灵敏性分析的应用可在文献 [3,157,167] 中找到.

最优化问题的灵敏度分析的其他各种方式由文献 [73] 给出.

7.4 应 用

7.4.1 变分不等式

变分不等式由 Stampacchia[204] 与 Lions 和 Stampacchia[14] 研究数学物理问题时引入. 在大批的文献中, 我们引用文献 [68], 此文以系统的方式研究了变分不等式在力学与物理中的应用, 对于数学方向的进展, 见文献 [49, 116]. 变分不等式解的灵敏性的最初的结果由 Mignot[152] 得到, 他用多面性的概念, 讨论了标准变分不等式 (即 $G(x) = x$) 的特殊情况, 假设相联系的二次形式是强凸的.

命题 5.2 与定理 5.9 的上 Lipschitz 稳定性结果推广了凸集 K 被有限多个等式与不等式约束参数化的有限维问题的结果[129]. 定理 5.5 与定理 5.10 的灵敏性分析是文献 [127] 当 $K = \mathbb{R}^n_+$ 的情况与文献 [171] 当 K 由有限多个等式与不等式约束参数化的情况的相应结果到无限维情况的推广. Barbet[18] 在强的规范条件下, 得到一类似的结果. 通过 “protoderivative” 来描述的同一本质性的结果由 Levy[137] 得到.

“广义方程” 的术语由 Robinson[77] 引入. 事实上, “变分不等式” 与 “广义方程” 是以不同的方式表示同一对象的词汇. 强正则性由 Robinson[175] 引入. 强正则性可推出一致二阶增长性质这一事实在一抽象的框架下给出证明[46]. 一致二阶增长条件在文献 [193] 中用于研究随机规划的统计估计的渐近性质时用到.

包含推论 5.41 的强正则性条件与二阶条件关系的讨论本质上是新的. 广义方程的强正则解的一阶展开由 Robinson[177] 得到. 对于非线性规划, 由命题 5.38 给出的强正则性的刻画在文献 [46] 中以直接的方式给出, 但也可由那一时期已知的结果推导出来. 非线性规划问题及具有多面集 K 的更一般的广义方程问题解的高阶展开源于文献 [46]. 有限维情况的推论在文献 [90] 中被综述. 线性互补问题的基本文献是 [63].

在非线性规划的情况, 强稳定性的概念由 Kojima[122] 引入. Kojima[122] 还给出非线性规划的强稳定性的代数刻画. 在局部解的情况, 这些代数条件等价于二阶充分性条件 (5.79)(如命题 5.37). 最近关于强稳定性结果的综述可参考文献 [121]. 由定理 5.34 给出的强稳定性与一致二阶增长性条件间的关系是新的. 非线性规划在线性无关条件下的强正则性与强稳定性间的等价关系可由 Kojima 与 Robinson 工作给出的一般性的刻画得到. 定理 5.35 将这一结果推广到 C^2 简约问题的情形.

在斜扰动下, 用广义Hesse 阵 (二阶次微分) 刻画局部极小点的强 Lipschitz 稳定性最近由 Poliquin 和 Rockafellar[167] 给出. "斜参数化" 一词受那篇文章的启发.

到 Hilbert 空间的凸子集 Φ 上的度量投影在一点 $x_0 \in \Phi$ 处的方向可微性由 Zarantonello[214] 证得. $\mathbb{R}^3$ 中凸集 Φ 使得相应的度量投影在点 $x_0 \notin \Phi$ 处非方向可微的例子由 Kruskal[124] 给出. $\mathbb{R}^2$ 中凸集的这样的例子在文献 [194] 中构造. 可以证明, 如果 Φ 是 (不必是凸的) 有限维 Hilbert(Euclid) 空间的子集, $x_0 \in \Phi$, 则如果 Φ 上在 x_0 处的内切锥与外切锥相等且为凸的, 那么相应的度量投影在 x_0 处是方向可微的[188]. 到 Hilbert 空间的凸的多面子集的度量投影的方向可微性由 Mignot[152] 与 Haraux[89] 证得. 到 Hilbert 空间的非凸子集 Φ 的度量投影在一点 $x_0 \in \Phi$ 处的方向可微性由 Shapiro[196] 研究. Shapiro[196] 证明, 对某一类 (不必是凸的) 集合 Φ, 相应的度量投影存在且在点 $x_0 \in \Phi$ 的邻域内是 Lipschitz 连续的. 这一类集合在文献 [167] 中被称为是 prox 正则的 (prox-regular), 在文献 [168] 中证明 prox 正则是度量投影这样性质的充分必要条件.

非光滑广义方程的灵敏度分析被 King 和 Rockafellar[117] 讨论.

7.4.2 非线性规划

前述章节的注记中对基本结果给出过评述. 然而, 我们注意到这样的事实, 尽管结果类似于第 3 章与第 4 章得到的结果, 主要的简化在于可以避免用多值函数的理论, 证明只用到隐函数定理. 这些技术已经被用到处理非线性规划问题的文献中.

Jittorntrum[110] 在起作用约束的梯度线性无关与充分性假设下, 得到非线性规划问题的最优解的局部唯一性与展开. 这一结果被 Janin [108] 推广, 在常秩假设之下, 确切地说, 在经验解处起作用的约束的任何子集在小的扰动下是保持不变的, 在 "半−强充分性条件" 下被 Bonnans[29] 推广, 在变分不等式的框架下被 Kyparisis[128] 推广. 这些方法在本书中未给出评述.

解的高阶展开式由 Bonnans 和 Sulem[46] 得到, 悬链问题的讨论取自作者的论文[45]. 高电压电子网络的讨论受 Bonnans[32] 工作的启发, 而交流电网络 (最优电力流量问题) 的类似理论由 Bonnans[33,35] 给出.

7.4.3 半定规划

半定规划问题被不同的作者冠以不同的名字被讨论若干年. 最近几年, 半定规

划成为研究热点, 应用的领域包括组合最优化、最优控制与统计学. 半定规划的一般性讨论可参看文献 [209] 与即将出版的关于半定规划的工具书[185].

矩阵的 Kronecker 积与相联系的向量代数的讨论可参看文献 [98]. 在文献 [98] 中有关于矩阵凸性的深入的讨论, 有矩阵凸性的一般性的充分性条件及较例 5.75 与例 5.76 更复杂的例子.

半定规划中如横截性条件的非退化性的方法由 Shapiro 和 Fan[202] 引入. 也被 Alizadeh, Haeberly 和 Overton[4] 与 Shapiro[199] 讨论. 文献 [4] 证明, 严格互补条件也是具有某种一般意义的. 解矩阵秩的一般的下界 (5.171) 被文献 [4,199] 讨论. 上界 (5.172) 来源于 Pataki[162].

SDP 问题的 Lagrange 乘子唯一性由 Shapiro[200] 讨论. 半定规划的二阶必要条件由 Shapiro[199] 证明. 相应的二阶充分性条件由 Bonnans, Cominetti 和 Shapiro[40] 证得. 半定规划的简约方法是新的. 定理 5.91 的结果来源于 Scheinberg[186].

7.4.4 半无限规划

Hettich 和 Kortanek[92] 的综述文章与 Goberna 和 López[83] 的书给出了半无限规划问题的各种例子及相关文献的列表. 命题 5.105 来源于 Borwein[48], 它推广了 Ben-Tal, Ben-Israel 和 Rossinger[21] 的工作, Helly 形式的定理被用于研究半无限规划问题的离散化性质.

半无限规划问题的一阶最优性条件是周知的[170]. (推广的)Mangasarian-Fromovitz 约束规范与 Lagrange 乘子集合的非空性与有界性的等价性 (定理 5.111) 是定理 3.9 的一般结果的具体结论. 对于半无限规划, 这在文献 [120] 中以直接的方式证得. 例 5.112 取自于 Klatte[118].

定理 5.114 中给出的 Lagrange 乘子唯一性的必要与充分性条件源于 Shapiro[200]. 例 5.115 取自于文献 [200].

半无限规划问题第一类型的二阶充分性条件通过所谓的简约化方法得到[22,93,212], 即相对应的下层最优化问题被设为是充分正则的, 使得最优值函数在集合 $\Delta(x_0)$ 的每一点的邻域内是二阶连续可微的, 此种情况下, 半无限规划问题可局部地简化为具有有限个光滑约束的非线性规划问题. 对半无限规划问题, 抽象形式的二阶最优性条件被 Kawasaki[114,115] 研究, 也可参看文献 [104,165,166].

7.5 最优控制

用于研究线性与半线性椭圆方程的 6.2 节的素材是经典的. 二阶椭圆方程的一个好文献是文献 [81], 关于 Sobolev 空间可参见文献 [1]. 偏微分方程的最优控制在文献 [138] 中以系统的方式进行研究. 状态控制问题的规范条件的检验在文献 [36]

中讨论. 二阶条件与二阶增长性的分析受 Bonnans[34] 工作的启发.

仅有某些约束是多面的情况的讨论可参看文献 [47]. 这包括关于控制的约束为有界, 具有有限多状态约束的情况, 在文献 [53,54] 中被讨论.

在变分计算中, "两范数" 方法是经典的思想[106], 一般用到的空间为 $L_2(0,T)$ 与 $L_\infty(0,T)$. 然而, 二阶条件的经典理论本质上处理的是无约束问题. 在假设推广了 Robinson[175] 的强正则性条件的强二阶条件成立的前提下, Malanowski[145~147] 讨论了到多面约束集合的推广. 这些问题的灵敏性分析被文献 [19,25,34,107,145,203] 以及其他文献讨论.

包含引理 6.41 的病态系统的最优控制问题在文献 [140] 中讨论. 到状态约束下问题的推广来自于 Bonnans 和 Cases[37]. 研究这样问题的另一技术, 基于所谓的 Fredholm 映射, 可参看文献 [82].

我们已经注意到关于多面性理论的文献. 容量理论由 Choquet[55] 给出. Clarke[58] 与 Clarke, Ledyaev, Stern 和 Wolenski[59] 讨论了具有非光滑数据的常微分方程组的最优控制问题.

Mignot[152] 和 Haraux[89] 讨论了偏微分方程组的变分不等式的最优控制, 还可参看文献 [27,153]. 正则性方法由 Barbu[20] 与 Tiba[206] 提出. Bergounioux 和 Mignot[26] 比较了处理问题的不同方法, 给出了最优性系统存在性的某些反例.

一个重要的悬而未解的问题是刻画状态约束问题(6.58) 的二阶增长性, 即使是线性状态方程的情形. 另一悬而未解的问题是这样问题的灵敏度分析. 对障碍问题的灵敏度分析如何推广到抛物变分不等式的情况也是一个悬而未解的问题. 另一个悬而未解的问题是此种情况下这些技术如何推广到奇异扰动情形[139].

参考文献

[1] Adams R A. *Sobolev spaces.* New York: Academic Press, 1975.

[2] Agmon S, Douglis A and Nirenberg L. Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions I. *Communications on Pure and Applied Mathematics*, 1959, 12: 623–727.

[3] Dontchev A L and Rockafellar R T. Characterization of strong regularity for variational inequalities over polyhedral convex sets. *SIAM Journal on Control and Optimization*, 1996, 6: 1087–1105.

[4] Alizadeh F, Haeberly J P and Overton M L. Complementarity and nondegeneracy in semidefinite programming. *Mathematical Programming, Series B*, 1997, 77: 111–128.

[5] Alt W. Stability of solutions to for a class of nonlinear cone constrained optimization problems. part 1: Basic theory. *Numerical Functional Analysis and Optimization*, 1989, 10: 1053–1064.

[6] Alt W. Local stability of solutions to differentiable optimization problems in Banach spaces. *Journal of Optimization Theory and Applications*, 1991, 70: 443–466.

[7] Anderson E J and Nash P. *Linear Programming in Infinite-Dimensional Spaces.* New York: Wiley, 1987.

[8] Anitescu M. Degenerate nonlinear programming with a quadratic growth constraint. *SIAM Journal on Optimization.* to appear.

[9] Attouch H. *Variational Convergence for Functions and Operators.* Boston: Pitman, 1984.

[10] Attouch H and Brézis H. Duality for the sum of convex functions in general Banach spaces//*Aspects of mathematics and its applications*, Barroso J A, 1986: 125–133.

[11] Attouch H and Wets R. A convergence theory for saddle functions. *Transactions of the American Mathematical Society*, 1981, 280: 1–41.

[12] Aubin J P. Comportement lipschitzien des solutions de problèmes de minimisation convexes. *Comptes Rendus de l'Académie des Sciences de Paris Série I*, 1982, 295: 235–238.

[13] Aubin J P and Ekeland I. *Applied nonlinear analysis.* New York: Wiley, 1984.

[14] Aubin J P and Frankowska H. *Set-valued analysis.* Boston: Birkhäuser, 1990.

[15] Auslender A and Cominetti R. First and second order sensitivity analysis of nonlinear programs under directional constraint qualification conditions. *Optimization*, 1990, 21: 351–363.

[16] Averbukh V I and Smolyanov O G. The various definitions of the derivative in linear topological spaces. *Russian Mathematical Surveys*, 1968, 23: 67–113.

[17] Bank B, Guddat J, Klatte D, Kummer B and Tammer K. *Nonlinear parametric optimization.* Berlin: Akademie Verlag, 1982.

[18] Barbet L. Analyse de sensibilité différentielle des solutions optimales d'inéquations variationnelles en dimension infinie. *Comptes Rendus de l'Académie des Sciences de Paris Série I*, 1992, 315: 1179–1182.

[19] Barbet L and Janin R. Analyse de sensibilité différentielle pour un problème d'optimisation en dimension infinie. *Comptes Rendus de l'Académie des Sciences de Paris Série I*, 1994, 318: 221–226.

[20] Barbu V. *Optimal control of variational inequalities*, volume 100 of *Research Notes in Mathematics.* Boston: Pitman, 1984.

[21] Ben-Tal A, Ben-Israel A and Rosinger E. A Helly-type theorem and semi-infinite programming//*Constructive Approaches to Mathematical models*, Coffman C V and Fix G J editors, 127–135. New York: Academic Press, 1979.

[22] Ben-Tal A, Teboulle M and Zowe J. Second order necessary optimality conditions for semi-infinite programming problems//*Semi-infinite programming*, volume 15 of *Lecture Notes in Control and Information Sciences.* Hettich R editor, 17–30. Berlin: Springer-Verlag, 1979.

[23] Ben-Tal A and Zowe J. Second order and related extremality conditions in nonlinear programming. *Journal of Optimization Theory and Applications*, 1980, 31: 143–165.

[24] Ben-Tal A and Zowe J. A unified theory of first and second order conditions for extremum problems in topological vector spaces. *Mathematical Programming Study*, 1982, 19: 39–76.

[25] Bergounioux M and Merabet N. Sensitivity analysis and optimal control of problems governed by semilinear parabolic equations. *Control and Cybernetics.* To appear.

[26] Bergounioux M and Mignot F. Control of variational inequalities and Lagrange multipliers. *Control, Optimization and Calculus of Variations (COCV)*, 2000, 5: 45–70.

[27] Bermúdez A and Saguez C. Optimal control of a Signorini problem. *SIAM Journal on Control and Optimization*, 1987, 25: 576–582.

[28] Bertsekas D P. *Constrained optimization and Lagrange multipliers methods.* New York: Academic Press, 1982.

[29] Bonnans J F. A semi-strong sufficiency condition for optimality in non convex programming and its connection to the perturbation problem. *Journal of Optimization Theory and Applications*, 1989, 60: 7–18.

[30] Bonnans J F. Directional derivatives of optimal solutions in smooth nonlinear programming. *Journal of Optimization Theory and Applications*, 1992, 73: 27–45.

[31] Bonnans J F. Exact penalization with a small nonsmooth term. *Revista de*

Matemáticas Aplicadas, 1996, 17: 37–45.

[32] Bonnans J F. Mathematical study of very high voltage power networks I: The optimal DC power flow problem. *SIAM Journal on Optimization*, 1997, 7: 979–990.

[33] Bonnans J F. Mathematical study of very high voltage power networks II: The AC power flow problem. *SIAM Journal on Applied Mathematics*, 1998, 58: 1547–1567.

[34] Bonnans J F. Second order analysis for control constrained optimal control problems of semilinear elliptic systems. *Journal of Applied Mathematics & Optimization*, 1998, 38: 303–325.

[35] Bonnans J F. Mathematical study of very high voltage power networks III: The optimal AC power flow problem. *Computational Optimization and Applications*, 2000, 16. to appear.

[36] Bonnans J F and Casas E. Contrôle de systèmes elliptiques semilinéaires comportant des contraintes sur l'état//*Nonlinear partial differential equations and their applications, Collège de France seminar vol. VII*, volume 166 of *Pitman Research Notes in Mathematics*. Brézis H and Lions J L, editors, 69–86. New York: Longman Scientific & Technical, 1988.

[37] Bonnans J F and Casas E. Optimal control of semilinear multistate systems with state constraints. *SIAM Journal on Control and Optimization*, 1989, 27: 446–455.

[38] Bonnans J F and Cominetti R. Perturbed optimization in Banach spaces I: A general theory based on a weak directional constraint qualification II: A theory based on a strong directional qualification, III: Semi-infinite optimization. *SIAM Journal on Control Optimization*, 1996, 34: 1151–1171, 1172–1189 and 1555–1567.

[39] Bonnans J F, Cominetti R and Shapiro A. Sensitivity analysis of optimization problems under abstract constraints. *Mathematics of Operations Research*, 1998, 23: 806–831.

[40] Bonnans J F, Cominetti R and Shapiro A. Second order optimality conditions based on parabolic second order tangent sets. *SIAM Journal on Optimization*, 1999, 9: 466–492.

[41] Bonnans J F and Ioffe A D. Quadratic growth and stability in convex programming problems with multiple solutions. *Journal of Convex Analysis*, 1995, 2: 41–57. (Special issue dedicated to R. T. Rockafellar).

[42] Bonnans J F and Ioffe A D. Second-order sufficiency and quadratic growth for non isolated minima. *Mathematics of Operations Research*, 1995, 20: 801–817.

[43] Bonnans J F, Ioffe A D and Shapiro A. Développement de solutions exactes et approchées en programmation non linéaire. *Comptes Rendus de l'Académie des Sciences de Paris, Série I*, 1992, 315: 119–123.

[44] Bonnans J F and Shapiro A. Nondegeneracy and quantitative stability of parameterized optimization problems with multiple solutions. *SIAM Journal on Optimization*,

1998, 8: 940–946.

[45] Bonnans J F and Shapiro A. Optimization problems with perturbations, A guided tour. *SIAM Review*, 1998, 40: 202–227.

[46] Bonnans J F and Sulem A. Pseudopower expansion of solutions of generalized equations and constrained optimization problems. *Mathematical Programming*, 1995, 70: 123–148.

[47] Bonnans J F and Zidani H. Optimal control problems with partially polyhedric constraints. *SIAM Journal on Control and Optimization*, 1999, 37: 1726–1741.

[48] Borwein J M. Direct theorems in semi-infinite convex programming. *Mathematical Programming*, 1981, 21: 301–318.

[49] Brézis H. Problèmes unilatéraux. *Journal de Mathématiques pures et appliquées*, 1972, 51: 1–168.

[50] Burke J V. Calmness and exact penalization. *SIAM Journal on Control and Optimization*, 1991, 29: 493–497.

[51] Burke J V and Poliquin R A. Optimality conditions for non-finite valued convex composite functions. *Mathematical Programming*, 1992, 57: 103–120.

[52] Burke J V and Ferris M C. Weak sharp minima in mathematical programming. *SIAM Journal on Control and Optimization*, 1993, 31: 1340–1359.

[53] Casas E and Tröltzsch F. Second order necessary optimality conditions for some state-constrained control problems of semilinear elliptic equations. *Journal of Applied Mathematics & Optimization*, 1999, 39: 211–227.

[54] Casas E, Tröltzsch F and Unger A. Second order sufficient optimality conditions for a nonlinear elliptic boundary control problem. *Zeitschrift für Analysis und ihre Anwendungen*, 1996, 15: 687–707.

[55] Choquet G. Theory of capacities. *Annales de l'Institut Fourier*, 1953–1954, 5: 131–292.

[56] Clarke F H. A new approach to Lagrange multipliers. *Mathematics of Operations Research*, 1976, 2: 165–174.

[57] Clarke F H. *Optimization and nonsmooth analysis*. New York: Wiley, 1983.

[58] Clarke F H. *Methods of dynamic and nonsmooth optimization*. volume 57 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia, 1989.

[59] Clarke F H, Ledyaev Yu S, Stern R J and Wolenski P. *Nonsmooth analysis and control theory*. volume 178 of *Graduate texts in Mathematics*. New York: Springer-Verlag, 1998.

[60] Cominetti R. Metric regularity, tangent sets and second order optimality conditions. *Journal of Applied Mathematics & Optimization*, 1990, 21: 265–287.

[61] Cominetti R and Penot J P. Tangent sets to unilateral convex set. *Comptes Rendus de l'Académie des Sciences de Paris, Série I*, 1995, 321: 1631–1636.

[62] Contesse B L. Une caractérisation complète des minima locaux en programmation mathématique. *Numerische Mathematik*, 1980, 34: 315–332.

[63] Cottle R W. Pang J-S and Stone R E. *The Linear Complementarity Problem.* New York: Academic Press, 1992.

[64] Danskin J M. *The Theory of Max-Min and Its Applications to Weapons Allocation Problems.* New York: Springer-Verlag, 1967.

[65] Dem'yanov V F and Malozemov V N. *Introduction to Minimax.* New York: Wiley, 1974.

[66] Dem'yanov V F and Pevny A B. Marginal values of problems of mathematical programming and minimax problems. *Vestnik LGU, Mat., Mekh., Astron.*, 1973, 4: 31–45.

[67] Dunford N and Schwartz J. *Linear operators, Vol I and II.* New York: Interscience, 1958, 1963.

[68] Duvaut G and Lions J L. *Les inéquations en mécanique et en physique.* Paris: Dunod, 1972.

[69] Ekeland I. Nonconvex minimization problems. *Bulletin of the American Mathematical Society*, 1979, 1(New Series): 443–474.

[70] Ekeland I and Temam R. *Convex analysis and variational problems*, volume 1 of *Studies in Mathematics and its Applications.* Amsterdam: North-Holland, 1976. French edition: Analyse convexe et problèmes variationnels. Paris: Dunod, 1974.

[71] Eremin I I. The penalty method in nonlinear programming. *Doklady Akad. Nauk SSSR*, 1966, 173: 459–462.

[72] Fiacco A V. *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming.* New York: Academic Press, 1983.

[73] Fiacco A V. *Mathematical programming with data perturbations*, volume 195 of *Lecture notes in pure and applied mathematics.* New York: Dekker M., 1997.

[74] Fiacco A V and McCormick G P. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques.* New York: Wiley, 1968.

[75] Frank M and Wolfe Ph. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 1956, 3: 95–110.

[76] Gauvin J. A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming. *Mathematical Programming*, 1977, 12: 136–138.

[77] Gauvin J and Dubeau F. Differential properties of the marginal function in mathematical programming. *Mathematical Programming Study*, 1982, 19: 101–119.

[78] Gauvin J and Janin R. Directional behaviour of optimal solutions in nonlinear mathematical programming. *Mathematics of Operations Research*, 1988, 13: 629–649.

[79] Gauvin J and Janin R. Directional Lipschitzian optimal solutions and directional derivative for the optimal value function in nonlinear mathematical programming.

Analyse non linéaire, 1989, 21: 305–324.

[80] Gauvin J and Tolle J W. Differential stability in nonlinear programming. *SIAM Journal on Control and Optimization*, 1977, 15: 294–311.

[81] Gilbarg D and Trudinger N S. *Elliptic partial differential equations of second order.* Berlin: Springer Verlag, 1983.

[82] Ginsburg B and Ioffe A D. The maximum principle in optimal control of systems governed by semilinear equations//*Nonsmooth Analysis and Geometric Methods in Deterministic Optimal Control*, volume 78 of *IMA Volumes in Mathematics and its Applications*, Mordukhovich B and Sussmann H, 81–110. Berlin: Springer Verlag, 1995.

[83] Goberna M A and Lopez M A. *Linear semi-infinite optimization*, volume 2 of *Wiley Series in Mathematical Methods in Practice.* Chichester: Wiley, 1998.

[84] Goldman A J and Tucker A W. Theory of linear programming. In *Linear inequalities and related systems*, volume 38 of *Annals of Mathematics Studies*, 53–97. Princeton, NJ: Princeton University Press, 1956.

[85] Gollan B. On the marginal function in nonlinear programming. *Mathematics of Operations Research*, 1984, 9: 208–221.

[86] Gol'shtein E G. *Theory of Convex Programming*, volume 36 of *Translations of Mathematical Monographs.* Providence: American Mathematical Society, 1972.

[87] Golubitsky M and Guillemin V. *Stable Mappings and their Singularities.* New York: Springer-Verlag, 1973.

[88] Graves L M. Some mapping theorems. *Duke Mathematical Journal*, 1950, 17: 111–114.

[89] Haraux A. How to differentiate the projection on a convex set in Hilbert space. Some applications to variational inequalities. *Journal Mathematical Society of Japan*, 1977, 29: 615–631.

[90] Harker P T and Pang J-S. Finite dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications. *Mathematical Programming, Series B*, 1990, 48: 161–220.

[91] Hestenes M R. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, 1969, 4: 303–320.

[92] Hettich R and Kortanek K O. Semi-infinite programming: theory methods and applications. *SIAM Review*, 1993, 35: 380–429.

[93] Hettich R P and Jongen H T. Semi-infinite programming: conditions of optimality and applications. In J. Stoer, editor, *Optimization Techniques.* Proc. 8th IFIP Conf. on Optimization Techniques, Würzburg. Part 2. New York: Springer-verlag, 1977.

[94] Hiriart-Urruty J B. Calculus rules on the approximate second-order directional derivative of a convex function. *SIAM Journal on Control and Optimization*, 1984, 22: 381–404.

[95] Hiriart-Urruty J B, Moussaoui M, Seeger A and Volle M. Subdifferential calculus without qualification conditions, using approximate subdifferentials: a survey. *Nonlinear Analysis, Theory, Methods and Applications*, 1995, 24: 1727–1754.

[96] Hiriart-Urruty J B and Seeger A. Règles de calcul sur le sous-différentiel second d'une function convexe. *Comptes Rendus de l'Académie des Sciences de Paris Série I*, 1987, 304: 259–262.

[97] Hoffman A. On approximate solutions of systems of linear inequalities. *Journal of Research of the National Bureau of Standards, Section B, Mathematical Sciences*, 1952, 49: 263–265.

[98] Horn R A and Jonhson C R. *Topics in Matrix Analysis*. Cambridge: Cambridge University Press, 1991.

[99] Ioffe A D. Codirectional compactness, metric regularity and subdifferential calculus//*Constructive, Experimental and Nonlinear Analysis*. Thera M, editor, Canadian Mathematical Society Proceedings Series. To appear.

[100] Ioffe A D. Metric regularity and subdifferential calculus. *Uspehi Mat. Nauk*. To appear.

[101] Ioffe A D. Necessary and sufficient conditions for a local minimum I: A reduction theorem and first order conditions, II: Conditions of Levitin-Miljutin-Osmolovskii type, III: Second order conditions and augmented duality. *SIAM Journal on Control Optimization*, 1979, 17: 245–250, 251–265 and 266-288.

[102] Ioffe A D. Regular points of Lipschitz functions. *Transactions of the American Mathematical Society*, 1979, 251: 61–69.

[103] Ioffe A D. Nonsmooth analysis: differential calculus of nondifferentiable mappings. *Transactions of the American Mathematical Society*, 1981, 266: 1–56.

[104] Ioffe A D. On some recent developments in the theory of second order optimality conditions.//*Optimization*, volume 1405 of *Lecture Notes in Mathematics*, Dolecki S, editor, pages 55–68. Berlin: Springer Verlag, 1989.

[105] Ioffe A D. Variational analysis of a composite function: A formula for the lower second order epiderivative. *Journal of Mathematical Analysis and Applications*, 1991, 160: 379–405.

[106] Ioffe A D and Tihomirov V M. *Theory of Extremal Problems*. Amsterdam, North-Holland Publishing Company, 1979. Russian Edition: Moscow: Nauka, 1974.

[107] Ito K and Kunish K. Sensitivity analysis of solutions to optimization problems in Hilbert spaces with applications to optimal control and estimation. *Journal of Differential Equations*, 1992, 99: 1–40.

[108] Janin R. Directional derivative of the marginal function in nonlinear programming. *Mathematical Programming*, 1984, 21: 110–126.

[109] Janin R and Gauvin J. Lipschitz-type stability in nonsmooth convex programs. *SIAM Journal on Control and Optimization*, 1999, 38: 124–137.

[110] Jittorntrum K. Solution point differentiability without strict complementarity in nonlinear programming. *Mathematical Programming*, 1984, 21: 127–138.

[111] John F. Extremum problems with inequalities as subsidiary conditions//*Studies and Essays, R. Courant anniversary volume*, 187–204. New York: Interscience, 1948.

[112] Jourani A and Thibault L. Coderivatives of multivalued mappings, locally compact cones and metric regularity. *Nonlinear Analysis, Applications and Methods*, 1999, 35: 925–945.

[113] Karush W. Minima of functions of several variables with inequalities as side constraints. Technical report, Department of Mathematics, University of Chicago, 1939. Master's thesis.

[114] Kawasaki H. An envelope-like effect of infinitely many inequality constraints on second order necessary conditions for minimization problems. *Mathematical Programming*, 1988, 41: 73–96.

[115] Kawasaki H. Second-order necessary and sufficient optimality conditions for minimizing a sup-type function. *Journal of Applied Mathematics & Optimization*, 1992, 26: 195–220.

[116] Kinderlehrer D and Stampacchia G. *An introduction to variational inequalities and their applications*. New York: Academic Press, 1980.

[117] King A J and Rockafellar R T. Sensitivity analysis for nonsmooth generalized equations. *Mathematical Programming*, 1992, 55: 193–212.

[118] Klatte D. On regularity and stability in semi-infinite optimization. *Set-Valued Analysis*, 1995, 3: 101–111.

[119] Klatte D. On quantitative stability for non-isolated minima. *Control and Cybernetics*, 1996, 23: 183–200.

[120] Klatte D and Henrion R. Regularity and stability in nonlinear semi-infinite optimization. *Nonconvex Optimization and Applications*, 1998, 25: 69–102.

[121] Klatte D and Kummer B. Strong stability in nonlinear programming revisited. *Journal of the Australian Mathematical Society Series B*, 1999, 40: 336–352.

[122] Kojima M. Strongly stable stationary solutions in nonlinear programming// *Analysis and computation of fixed points*, Robinson S M, editor, 93–138. New York: Academic Press, 1980.

[123] Kretschmer K S. Programs in paired spaces. *Canadian Journal Math.*, 1961, 13: 221–238.

[124] Kruskal J B. Two convex counterexamples: a discontinuous envelope function and a non-differentiable nearest point mapping. *Proceedings of the American Mathematical Society*, 1969, 23: 679–703.

[125] Kuhn H W and Tucker A W. Non-linear programming//*Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Neyman J editor,

481–493. Berkeley: University of California Press, 1951.

[126] Kurcyusz S. On the existence and nonexistence of Lagrange multipliers in Banach spaces. *Journal of Optimization Theory and Applications*, 1976, 20: 81–110.

[127] Kyparisis J. Uniqueness and differentiability of solutions of parametric nonlinear complementarity problems. *Mathematical Programming*, 1986, 36: 105–113.

[128] Kyparisis J. Sensitivity analysis for nonlinear programs and variational inequalities with nonunique multipliers. *Mathematics of Operations Research*, 1990, 15: 286–298.

[129] Kyparisis J. Parametric variational inequalities with nultivalued solution sets. *Mathematics of Operations Research*, 1992, 17: 341–364.

[130] Lagrange J L. *Théorie des fonctions analytiques*. Paris, 1813//*Oeuvres*. Paris: Gauthier-Villars, 1867–1892.

[131] Lempio F and Maurer H. Differential stability in infinite-dimensional nonlinear programming. *Journal of Applied Mathematics & Optimization*, 1980, 6: 139–152.

[132] Levin V L. Application of a theorem of E. Helly in convex programming, problems of best approximation and related topics. *Mat. Sbornik*, 1969, 79: 250–263.

[133] Levitin E S. On differential properties of the optimal value of parametric problems of mathematical programming. *Dokl. Akad. Nauk SSSR*, 1975, 224: 1354–1358.

[134] Levitin E S. Differentiability with respect to a parameter of the optimal value in parametric problems of mathematical programming. *Kibernetika*, 1976, 1: 44–59.

[135] Levitin E S. *Perturbation theory in mathematical programming and its applications*. New York: Wiley, 1994.

[136] Levitin E S, Miljutin A A and Osmolovskii N P. On conditions for a local minimum in problem with constraints//*Mathematical economics and functional analysis*. Mitjagin B S editor. Moscow: Nauka, 1974. In Russian.

[137] Levy A B. Sensitivity of solutions to variational inequalities on Banach spaces. *SIAM Journal Control and Optimization*, 1999, 38: 50–60.

[138] Lions J L. *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles*. Paris: Dunod, 1968.

[139] Lions J L. *Perturbations singulières dans les problèmes aux limites et en contrôle optimal*, volume 323 of *Lecture Notes in Mathematics*. Berlin: Springer-Verlag, 1973.

[140] Lions J L. *Contrôle des systèmes distribués singuliers*. Paris: Dunod, 1983.

[141] Lions J L and Stampacchia G. Variational inequalities. *Communications in Pure and Applied Mathematics*, 1967, 20: 493–519.

[142] Luo Z-Q and Zhang S. On extensions of the Frank-Wolfe theorems. *Computational Optimization and Applications*, 1999, 13: 87–110.

[143] Lyusternik L. Conditional extrema of functions. *Math. USSR-Sb*, 1934, 41: 390–401.

[144] Majthay A. Optimality conditions for quadratic programming. *Mathematical Programming*, 1971, 1: 359–365.

[145] Malanowski K. Second order conditions and constraint qualifications in stability and sensitivity analysis of solutions to optimization problems in Hilbert spaces. *Journal of Applied Mathematics & Optimization*, 1992, 25: 51–79.

[146] Malanowski K. Two-norm approach in stability and sensitivity analysis of optimization and optimal control problems. *Advances in Mathematical Sciences and Applications*, 1993, 2: 397–443.

[147] Malanowski K. Stability and sensitivity of solutions to nonlinear optimal control problems. *Journal of Applied Mathematics & Optimization*, 1995, 32: 111–141.

[148] Mangasarian O. Locally unique solutions of quadratic programs, linear and nonlinear complementarity problems. *Mathematical Programming*, 1980, 19: 200–212.

[149] Mangasarian O and Fromovitz S. The Fritz-John necessary optimality conditions in the presence of equality and inequality constraints. *Journal of Mathematical Analysis and Applications*, 1967, 7: 37–47.

[150] Mangasarian O L and Shiau T H. Lipschitz continuity of solutions of linear inequalities, programs and complementarity problems. *SIAM Journal on Control and Optimization*, 1987, 25: 583–595.

[151] Maurer H and Zowe J. First and second-order necessary and sufficient optimality conditions for infinite-dimensional programming problems. *Mathematical Programming*, 1979, 16: 98–110.

[152] Mignot F. Contrôle dans les inéquations variationnelles elliptiques. *Journal of Functional Analysis*, 1976, 22: 130–185.

[153] Mignot F and Puel J P. Optimal control in some variational inequalities. *SIAM Journal on Control and Optimization*, 1984, 22: 466–476.

[154] Mordukhovich B S. Maximum principle in problems of time optimal control with nonsmooth constraints. *Journal of Applied Mathematics and Mechanics*, 1976, 40: 960–969.

[155] Mordukhovich B S. Metric approximations and necessary optimality conditions for general classes of nonsmooth extremal problem. *Soviet Mathematics Doklady*, 1980, 22: 526–530.

[156] Mordukhovich B S. Complete characterization of openness, metric regularity, and Lipschitzian properties of multifunctions. *Transactions of the American Mathematical Society*, 1993, 340: 1–35.

[157] Mordukhovich B S. Coderivatives of set-valued mappings: calculus and applications. *Nonlinear Analysis, Theory, Methods and Applications*, 1997, 30: 3059–3070.

[158] Mordukhovich B S and Shao Y. Stability of set-valued mappings in infinite dimensions: point criteria and applications. *SIAM Journal on Control and Optimization*, 1997, 35: 285–314.

[159] Nashed M Z. Differentiability and related properties of nonlinear operators: Some

aspects of the role of differentials in nonlinear functional analysis.//*Nonlinear Functional Analysis and Applications*. Rall L B editor, 103–309. New York: Academic Press, 1971.

[160] Nečas J. *Les méthodes directes en théorie des équations elliptiques*. Paris: Masson, 1967.

[161] Páles Zs and Zeidan V M. Nonsmooth optimum problems with constraints. *SIAM Journal on Control and Optimization*, 1994, 32: 1476–1502.

[162] Pataki G. On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues. *Mathematics of Operations Research*, 1998, 23: 339–358.

[163] Penot J P. A characterization of tangential regularity. *Nonlinear Analysis, Theory, Methods and Applications*, 1981, 5: 625–643.

[164] Penot J P. On regularity conditions in mathematical programming. *Mathematical Programming Study*, 1982, 19: 167–199.

[165] Penot J P. Optimality conditions for minimax problems, semi-infinite programming problems and their relatives. Technical Report 92/16. UPRA. Laboratoire de Mathématiques Appliquées. Avenue de l'Université. 64000 Pau. France, 1992.

[166] Penot J P. Optimality conditions in mathematical programming and composite optimization. *Mathematical Programming*, 1994, 67: 225–245.

[167] Poliquin R A and Rockafellar R T. Tilt stability of a local minimum. *SIAM Journal on Optimization*, 1998, 8: 287–299.

[168] Poliquin R A, Rockafellar R T and Thibault L. Local differentiability of distance functions. *Transactions of the American Mathematicans Society*. To appear.

[169] Powell M J D. A method for nonlinear constraints in minimization problems.//*Optimization*. Fletcher R editor, 283–298. New York: Academic Press, 1969.

[170] Pshenichnyi B N. *Necessary Conditions for an Extremum*. New York: Marcel Dekker, 1971.

[171] Qiu Y and Magnanti T L. Sensitivity analysis for variational inequalities. *Mathematics of Operations Research*, 1992, 17: 61–76.

[172] Robinson S M. First order conditions for general nonlinear optimization. *SIAM Journal on Applied Mathematics*, 1976, 30: 597–607.

[173] Robinson S M. Regularity and stability for convex multivalued functions. *Mathematics of Operations Research*, 1976, 1: 130–143.

[174] Robinson S M. Stability theorems for systems of inequalities, part II: differentiable nonlinear systems. *SIAM Journal on Numerical Analysis*, 1976, 13: 497–513.

[175] Robinson S M. Strongly regular generalized equations. *Mathematics of Operations Research*, 1980, 5: 43–62.

[176] Robinson S M. Some continuity properties of polyhedral multifunctions. *Mathematical Programming Study*, 1981, 14: 206–214.

[177] Robinson S M. Generalized equations and their solutions. part II: Applications to nonlinear programming. *Mathematical Programming Study*, 1982, 19: 200–221.

[178] Robinson S M. Local structure of feasible sets in nonlinear programming, part II: Nondegeneracy. *Mathematical Programming Study*, 1984, 22: 217–230.

[179] Rockafellar R T. *Convex Analysis.* New Jersey: Princeton University Press, 1970.

[180] Rockafellar R T. *Conjugate Duality and Optimization.* Number 16 in Regional Conference Series in Applied Mathematics. SIAM. Philadelphia, 1974.

[181] Rockafellar R T. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1976, 1: 97–116.

[182] Rockafellar R T. First and second-order epi-differentiability in nonlinear programming. *Transactions of the American Mathematical Society*, 1988, 307: 75–108.

[183] Rockafellar R T. Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives. *Mathematics of Operations Research*, 1989, 14: 462–484.

[184] Rockafellar R T and Wets R J-B. *Variational Analysis.* New York: Springer-Verlag, 1997.

[185] Saigal R, Vandenberghe L and Wolkowicz H editors. *Semidefinite Programming and Applications Handbook.* Boston: Kluwer Academic Publishers. To appear.

[186] Scheinberg K. Parametric linear semidefinite programming// Saigal R, Vandenberghe L and Wolkowicz H editors. *Semidefinite Programming and Applications Handbook.* Boston: Kluwer Academic Publishers. To appear.

[187] Shapiro A. Second-order sensitivity analysis and asymptotic theory of parametrized, nonlinear programs. *Math. Programming*, 1985, 33: 280–299.

[188] Shapiro A. On differentiability of metric projections in $\mathbb{R}^n$, 1: Boundary case. *Proceedings of the American Mathematical Society*, 1987, 99: 123–128.

[189] Shapiro A. Perturbation theory of nonlinear programs when the set of optimal solutions is not a singleton. *Journal of Applied Mathematics & Optimization*, 1988, 18: 215–229.

[190] Shapiro A. Sensitivity analysis of nonlinear programs and differentiability properties of metric projections. *SIAM Journal on Control and Optimization*, 1988, 26: 628–645.

[191] Shapiro A. On concepts of directional differentiability. *Journal of Optimization Theory and Applications*, 1990, 66: 477–487.

[192] Shapiro A. Perturbation analysis of optimization problems in Banach spaces. *Numerical Functional Analysis and Optimization*, 1992, 13: 97–116.

[193] Shapiro A. Asymptotic behavior of optimal solutions in stochastic programming. *Mathematics of Operations Research*, 1993, 18: 829–845.

[194] Shapiro A. Directionally nondifferentiable metric projection. *Journal of Optimization*

Theory and Applications, 1994, 81: 203–204.

[195] Shapiro A. On Lipschitzian stability of optimal solutions of parametrized semi-infinite programs. *Mathematics of Operations Research*, 1994, 19: 743–752.

[196] Shapiro A. Sensitivity analysis of parametrized programs via generalized equations. *SIAM Journal on Control and Optimization*, 1994, 32: 553–571.

[197] Shapiro A. Directional defferentiability of the optimal value function in convex semi-infinite programming. *Mathematical programming, Series A*, 1995, 70: 149–157.

[198] Shapiro A. A variational principle and its applications//*Proceedings of the Conference on "Parametric Optimization and Related Topics I"*. Guddat J, Jongen H Th, Nozicka F, Still G and Twilt F editors. Approximation and Optimization, 341–357. Frankfurt: Peter Lang Verlag, 1996.

[199] Shapiro A. First and second order analysis of nonlinear semidefinite programs. *Mathematical Programming Series B*, 1997, 77: 301–320.

[200] Shapiro A. On uniqueness of Lagrange multipliers in optimization problems subject to cone constraints. *SIAM Journal on Optimization*, 1997, 7: 508–518.

[201] Shapiro A and Bonnans J F. Sensitivity analysis of parametrized programs under cone constraints. *SIAM Journal on Control and Optimization*, 1992, 30: 1409–1422.

[202] Shapiro A and Fan M K H. On eigenvalue optimization. *SIAM Journal on Optimization*, 1995, 5: 552–569.

[203] Sokolowski J. Sensitivity analysis of control constrained optimal control problems for distributed parameter systems. *SIAM Journal on Control and Optimization*, 1987, 25: 1542–1556.

[204] Stampacchia G. Formes bilinéaires coercitives sur les ensembles convexes. *Comptes Rendus de l'Académie des Sciences de Paris Série I*, 1964, 258: 4413–4416.

[205] Studniarski M and Ward D E. Weak sharp minima: characterizations and sufficient conditions. *SIAM Journal on Control and Optimization*, 1999, 38: 219–236.

[206] Tiba D. *Optimal control of nonsmooth distributed parameter systems*, volume 1459 of *Lecture notes in mathematics*. Berlin: Springer-Verlag, 1990.

[207] Ursescu C. Multifunctions with convex closed graph. *Czechoslovak Mathematical Journal*, 1975, 25: 438–441.

[208] Valadier M. Sous-différentiels d'une borne supérieure et d'une somme continue de fonctions convexes. *Comptes Rendus de l'Académie des Sciences de Paris Série A*, 1969, 268: 39–42.

[209] Vandenberghe L and Boyd S. Positive definite programming. *SIAM Review*, 1996, 38: 49–95.

[210] Ward D E. Characterization of strict local minima and necessary conditions for weak sharp minima. *Journal of Optimization Theory and Applications*, 1994, 80: 551–571.

[211] Ward D E. Sufficient conditions for weak sharp minima of order two and directional

derivatives of value function//*Mathematical programming with data perturbations*, volume 195 of *Lecture notes in pure and applied mathematics*. Fiacco A V editor, New York: M. Dekker, 1997.

[212] Wetterling W. Definitheitsbedingungen für relative Extrema bei Optimierungs-und Approximationsaufgaben. *Numerische Mathematik*, 1970, 15: 122–136.

[213] Yosida K. *Functional Analysis*. Berlin: Springer-Verlag, 1965.

[214] Zarantonello E H. Projections on convex sets in Hilbert space and spectral theory I and II. In *Contributions to Nonlinear Functional Analysis*, 237–424. New York: Academic Press, 1971.

[215] Zowe J and Kurcyusz S. Regularity and stability for the mathematical programming problem in Banach spaces. *Journal of Applied Mathematics & Optimization*, 1979, 5: 49–62.

索　引

E

F

G

H

J

K

L

N

P

Q

R

S

T

W

X

Y

Z

其他

(O-3035.0101)


ISBN 978-7-03-020429-5

销售分类建议：高等数学

定 价：88.00 元